Coversheet: Countering violent extremism online – changes to censorship legislation to better protect New Zealanders from online harm

Advising agencies	*Department of Internal Affairs*
Decision sought	*Cabinet policy approval*
Proposing Ministers	*Minister of Internal Affairs*

Summary: Problem and Proposed Approach

Problem Definition

What problem or opportunity does this proposal seek to address? Why is Government intervention required?

Gaps in the Films, Videos and Publications Classification (FVPC) Act 1993 were highlighted in the wake of the 15 March 2019 Christchurch Terror Attacks (the Terror Attacks). While government agencies, internet service providers (ISPs) and online content hosts worked collaboratively to respond to the Terror Attacks, the situation showed limitations in the FVPC Act's ability to enable a rapid response to prevent harm from exposure to objectionable content online.

Collectively, the identified limitations restrict the ability of Government to:

- remove objectionable content swiftly to prevent the risk of harm;
- deter the spread of objectionable content online; and
- act against those who create objectionable content and disseminate it online.

The following terms used throughout this paper are defined in the glossary (Section 8):

- harm;
- ISPs and online content hosts;
- objectionable; and
- violent extremism.

Proposed Approach

How will Government intervention work to bring about the desired change? How is this the best option?

There are opportunities right now, ahead of a planned review of the broader media content regulation system, to improve the regulatory framework of the FVPC Act to protect people in New Zealand from online harm and increase online safety. The proposed suite of legislative changes is considered the best and preferred option because it will facilitate:

- faster decision making about objectionable online content, including livestreaming, so

that decisions can be made swiftly and communicated to the public and industry; and
- introduction of improved mechanisms (including corresponding penalties for non-compliance) to reduce harm from exposure to objectionable online content.

Section B: Summary Impacts: Benefits and costs

Who are the main expected beneficiaries and what is the nature of the expected benefit?

Adjustments to regulatory settings and related processes will reduce the potential of New Zealanders being harmed from exposure to objectionable content online.

Victims of violent extremist or terrorist attacks and their families will be better protected from re-victimisation (i.e. having others view footage of their experience, or being unintentionally exposed to this content, causing further distress or trauma).

'Innocent bystanders' will be better protected from harm associated with exposure to objectionable content on their mainstream social media feeds.

Regulators and agency personnel will be better equipped to enforce the removal of objectionable content from websites, reducing harm to users. They will also be able to undertake further work and consult with relevant parties to explore options for blocking objectionable content online.

ISPs and online content hosts will benefit from the increased legal certainty and guidance on how to respond to objectionable content online through clearer legislative provisions.

Where do the costs fall?

Regulators including the Department of Internal Affairs and the Office of Film and Literature Classification will incur some limited additional costs as they will be undertaking more work than before to classify content and, in cases of non-compliance, to issue take-down notices, which may require establishing new operational processes.

Some online content hosts and social media companies may incur some costs in complying with take-down notices (i.e. removing content) under the proposed changes.

The minority of online content hosts that do not comply with legislative provisions for objectionable content online would be subject to penalties and offences.

All parties would incur resource costs in further work to explore and investigate the possibility of establishing filtering and blocking mechanisms, building on engagement and collaboration to date.

What are the likely risks and unintended impacts, how significant are they and how will they be minimised or mitigated?

In relation to the proposed changes we have identified several risks as follows:

- new decision-making powers around classification, interception and removal of online content, or new offences, are applied in situations other than intended or envisaged;

- ISPs and online content hosts react negatively to a more restrictive regulation of online content;
- ISPs and online content hosts become more risk averse and restrict content that is not objectionable;
- take-down powers are ineffective or have adverse impacts;
- offshore-domiciled companies choose to escape liability introduced by new or clarified penalties and offences by disregarding any laid charges;
- civil society views the changes as unduly limiting freedom of expression;
- the proposed changes are viewed by stakeholders as premature and more appropriately considered as part of the broader media regulation review;
- the proposed changes are inconsistent with the partnership approach agreed through the Christchurch Call[1]; and
- mechanisms for blocking objectionable content are seen as ineffective, easy to circumvent, or an infringement on freedoms of expression.

We consider we can mitigate these risks appropriately (refer to section 5.3 for these mitigations).

Identify any significant incompatibility with the Government's 'Expectations for the design of regulatory systems'.
None identified (refer to section 5.4 for more detail).

Section C: Evidence certainty and quality assurance

Agency rating of evidence certainty?
DIA has a high level of evidence certainty of the virality and related harm and negative consequences from violent extremist content, drawn from DIA, OFLC, ISP and online content hosts' incident reporting and experiences in dealing with the Terror Attacks. Recent information provided by online content hosts demonstrates the continued threat posed by objectionable content – the Terror Attacks video continues to be uploaded to online platforms and removed by online content hosts. DIA also has a high level of evidence certainty of the harm and negative consequences of child sexual exploitation material (CSEM), from the DIA Digital Safety Group.

[1] The Christchurch Call was a summit co-hosted by New Zealand's Prime Minister, Jacinda Ardern and French President, Emmanuel Macron, in Paris on 15 May 2019. The Call outlines collective, voluntary commitments from Governments and online service providers intended to address the issue of terrorist and violent extremist content online and to prevent the abuse of the internet as occurred in and after the Terror Attacks.

To be completed by quality assurers:

Quality Assurance Reviewing Agency:
Joint panel – Treasury/DIA
Quality Assurance Assessment:
Partially meets
Reviewer Comments and Recommendations:
The Department's Regulatory Impact Analysis (RIA) panel (the panel) has reviewed the 'Countering Violent Extremism' RIA (the RIA) in accordance with the quality assurance criteria set out in the CabGuide. This was a joint review by the Department and the Treasury. The panel members for this review were: • John Sutton, Principal Policy Analyst (Chair) • Rowan Burns, Senior Policy Analyst (Policy member) • Killian Destremau, Senior Analyst, The Treasury (External member) • Harry Boam, Policy Analyst (Secretariat member) The panel considers that the information and analysis summarised in the RIA partially meets the quality assurance criteria. The panel acknowledges the high level of public and political interest in these proposals. The Government's desire for a quick response to the Christchurch terror attacks has constrained both the scope and the time available for the analysis described in the RIA. Similarly, the focus on legislation administered by the Department has limited the policy and regulatory options available for assessment. The RIA is complete and includes all necessary information. Consultation with stakeholders was focussed primarily on the regulatory proposals, due to the time and scope constraints on the analysis. Discussions from this consultation informed refinement of the proposals and a commitment to release an exposure draft of the amendment Bill. The RIA would have been more convincing if it had clearly and concisely set out the contextual constraints and how these have impacted the scope and analysis of the defined problem and the identification and assessment of policy options. Without this clarity, the descriptions and justifications of the different elements of the preferred option are uneven and not fully convincing. The analysis of options could also have presented a clearer picture of the relative merits of urgent amendments to address regulatory gaps in the Films, Videos and Publications Classification Act 1993 against reliance on voluntary cooperation with industry stakeholders pending the planned broader review of media regulation. Finally, the RIA indicates that some detailed aspects of the proposed mechanisms are to be designed in subsequent stages, including through collaboration with industry. It acknowledges design and implementation challenges and sets out mitigation strategies which will be important in this critical dimension of the proposals.

Impact Statement: Countering violent extremism online – changes to censorship legislation to better protect New Zealanders from online harm

Section 1: General information

Purpose
The Department of Internal Affairs (DIA) is solely responsible for the analysis and advice set out in this Regulatory Impact Assessment (RIA), except as otherwise explicitly indicated. This analysis and advice has been produced to inform final Cabinet decisions to proceed with legislative and policy changes.
Key Limitations or Constraints on Analysis
Implementing changes to respond to the 15 March 2019 Terror Attacks (the Terror Attacks) is a priority for the Government and the policy development process has been undertaken under significant time pressure. In September 2019, Cabinet directed DIA to address harm from exposure to online content that falls within its regulatory jurisdiction, i.e. the Films, Videos and Publications Classification Act 1993 (FVPC Act). The scope of possible responses has consequently been limited to matters that can be addressed through the FVPC Act. The proposals in this RIA are part of the DIA-led workstream to counter violent extremism online, under the broader government work programme to counter violent extremism.[2] Cabinet requested a report-back on detailed policy proposals in December 2019 and required policy proposals to be informed by targeted stakeholder engagement. This timeline placed significant time pressure on the development of policy proposals and meant that they were developed iteratively and in parallel to stakeholder engagement. While time pressures limited the overall extent of the upfront policy work pre-engagement, the proposals have evolved to best align with stakeholder views over this period. Given the time constraints, engagement to inform policy development was limited to specific affected stakeholder groups. However, there will be wider public consultation as part of the Select Committee process in early 2020. Time constraints limited the extent to which a detailed analysis of feedback could be undertaken. Key themes emerged nonetheless that informed the development of the proposals. Before and throughout consultation on these proposals, civil society and community groups have been clear they expect the government to make immediate changes to protect people

[2] The proposals complement other domestic and international efforts to combat violent extremism, including: increased operational capacity in both DIA and OFLC to identify and assess violent extremist content online [CAB-10-MIN-0498 refers]; [Redacted under S9(2)(f)(iv)], and a cyber security strategy which sets out New Zealand's cyber policy positions, including the importance of maintaining a free, open and secure internet, and will guide any action to address online harms (refer to https://dpmc.govt.nz/sites/default/files/2019-07/Cyber%20Security%20Strategy.pdf).

from exposure to objectionable content online.

A “first principles response” to how social media companies and the wider internet are regulated could be an alternative approach. However, that would be an endeavour on a much greater scale. Agencies and Ministers will consider broader issues relating to the media regulatory system and its fitness for purpose in a contemporary, digital world as part of a broader media regulation review planned to be initiated in 2020.

This RIA relies on credible assumptions about both the status quo and the expected costs and benefits of the proposed changes (for example the harm associated with exposure to objectionable content online, the likely behaviour of actors involved and the increased legal certainty for Internet Service Providers and online content hosts). While most of the quantitative data on the nature of harm that we have used relate specifically to the single event of the Terror Attacks, we have also drawn on the experience of similar organisational entities in Australia and the volume of content they deal with. Costs are difficult to estimate precisely and will be informed by further and ongoing stakeholder input. It is challenging to accurately predict broad benefits, which will be determined by many interacting factors.

Responsible Manager (signature and date):

Raj Krishnan
General Manager Policy
Policy Regulation and Communities
Department of Internal Affairs

Section 2: Problem definition and objectives

2.1 What is the context within which action is proposed?

Objectionable content causes significant harm to those who view it

DIA maintains the FVPC Act, which regulates objectionable content (including violent extremism and terrorism) and sets out the offences against the Act as well as penalties for breaching New Zealand's classification regime. For the purposes of the FVPC Act, a publication is objectionable if it describes, depicts, expresses, or otherwise deals with matters such as sex, horror, crime, cruelty, or violence in such a manner that the availability of the publication is likely to be injurious to the public good.

There is increasing evidence about the negative psychological effects that individuals may experience due to online exposure to mass violence and acts of violent extremism and in the media.[3] Online violent extremist content causes harm in several ways:

- *Simply viewing this content can cause harm*. There are negative effects from exposure to mass violence and terrorism online.
- *Re-victimisation*. Having others see footage of your experience or being unintentionally exposed to the content can all cause further distress or trauma.
- *Radicalisation*. Violent extremist content can be used to promote the ideologies and actions of violent extremists and to influence others to incite similar acts.
- *Learning how to commit terrorist acts.* Online dissemination of acts of violent extremism can provide examples to others of the skills necessary to plan and carry out similar acts.

The Terror Attacks highlighted limitations of the FVPC Act to enable a rapid response to prevent and combat harm from objectionable content online

The FVPC Act contains mechanisms to deter people from creating or sharing objectionable content, to allow authorities to investigate those who do and to prosecute them where appropriate. For example, the FVPC Act allows defined inspectors of publications to seize illegal content, such as copies of child sexual exploitation material. The FVPC Act also sets out criminal penalties for offences under its ambit.

The Terror Attacks, and the virality of the online video of the Terror Attacks, highlighted gaps in the FVPC Act. These gaps are discussed in detail in section 2.3 below.

Updating specific areas of the FVPC Act now will give effect to the Government's Christchurch Call[4] commitment to 'ensure effective enforcement of applicable laws that prohibit the production or dissemination of terrorist and violent extremist content, in a manner consistent with the rule of law and international human rights law, including freedom of expression'.

---

[3] "Media exposure to mass violence events can fuel a cycle of distress", Rebecca R. Thompson, Nicholas M. Jones, E. Alison Holman and Roxane Cohen Silver, https://advances.sciencemag.org/content/5/4/eaav3502?rss=1

[4] "The Christchurch Call is a commitment by Governments and tech companies to eliminate terrorist and violent extremist content online. It rests on the conviction that a free, open and secure internet offers extraordinary benefits to society. Respect for freedom of expression is fundamental. However, no one has the right to create and share terrorist and violent extremist content online. See https://www.christchurchcall.com/

It is also important to note that some ISPs and online content hosts such as Facebook have expressed a desire for additional legal certainty on how the censorship system should operate in the contemporary digital environment, particularly in the face of viral violent extremist content, to provide a better basis for action than interpretations by individual private companies.

2.2 What regulatory system, or systems, are already in place?

The Films, Videos, and Publications Classification Act 1993 (the FVPC Act) is the main legislative framework

The FVPC Act contains the definition of objectionable content (which includes violent extremist and terrorist content), establishes the role of the Chief Censor, and sets out the offences and penalties for breaching content labelling requirements and offences and penalties relating to objectionable content. The OFLC (an independent Crown entity) classifies publications (films, books or computer files), and provides information about classification decisions and New Zealand's classification regime. DIA administers the Act and has a regulatory role under the FVPC Act where the Secretary of Internal Affairs appoints inspectors of publications (who are invested with powers to seize objectionable publications). The power to 'seize publications' illustrates that the FVPC Act is still largely based on and intended for a physical analogue media. However, the definition of publication (contained in the FVPC Act) also includes 'digital content' and can be applied to digital media online.

Other statutes relevant to controlling potentially harmful content online

Various statutes are relevant to the consideration of regulating media content in New Zealand. These statutes are not administered by DIA and include the following:

- **The Harmful Digital Communications Act 2015 (the HDC Act)** addresses digital communications that cause a harm to 'a victim'. The HDC Act does not apply to communications causing harm to groups of people. It currently provides a 'safe harbour' for content hosts, so long as they follow a set of requirements if someone makes a complaint. This 'safe harbour' was designed to prevent companies from being unwittingly exposed to liability from the actions of others.
- **The Terrorism Suppression Act 2002 (the TS Act)** includes an offence relating to recruiting members of terrorist groups, which can occur online.
- **The Human Rights Act 1993 (the HR Act)** establishes the Human Rights Commission which addresses behaviour that infringes upon human rights, including online content.
- **The Broadcasting Act 1989** provides for the maintenance of programme standards in broadcasting in New Zealand.

A broader systemic review of the regulatory framework is planned

The FVPC Act and the Broadcasting Act 1989 were drafted over 25 years ago, before the rise of sophisticated digital technologies, widespread use of internet, and the emergence of social media. The current media regulatory framework is no longer fit for purpose, and fragmented. Ensuring an appropriate regulatory response to the increasing convergence of telecommunications, information technology, media and entertainment sectors in New Zealand has been a cause for concern over the past decade.

Cabinet has directed the Minister of Internal Affairs and the Minister of Broadcasting,

Communications and Digital Media to scope a review to modernise New Zealand's media content regulatory system. Ministers have been invited to report to Cabinet in early 2020 with proposed terms of reference, engagement approach and timeline. This review is expected to take 18-24 months and will be undertaken jointly by DIA and the Ministry for Culture and Heritage.

2.3 What is the policy problem or opportunity?

There is substantive case law and precedent for the classification of the content of 'traditional' analogue media (e.g. books, films, posters, video games) under the FVPC Act. However, the extent to which the FVPC Act can be applied to digital online content is incomplete, as was highlighted by the Terror Attacks (the Terror Attacks). For example, while inspectors can seize physical objectionable material, there is no equivalent power applying to online content. While government agencies, ISPs and online content hosts worked collaboratively to respond as quickly as possible, the situation showed limitations in the FVPC Act's ability to respond swiftly and effectively to the sudden appearance and viral distribution of objectionable content online.

The FVPC Act also does not enable the Government to deal with ISPs and online content hosts that facilitate the creation and sharing of objectionable content through their online infrastructure and platforms. The Terror Attacks Video and Manifesto illustrated how fast objectionable material can spread online, and the fact that processes based on physical content are not fast enough to deal with the risks posed by online objectionable content.

The suite of proposals in this paper aim to better prevent and combat harms caused by objectionable content online. The proposals fit within the framework of the existing legislation DIA administers and will be progressed alongside an evolving partnership with online content hosts and civil society, arising from the Christchurch Call.

Six aspects of current legislation limit an effective response to objectionable content online

First, the act of **livestreaming objectionable content is not explicitly an offence against the FVPC Act**. Livestreaming – i.e. the act of using internet infrastructure to broadcast events as they occur in real time – is not covered by current legislation. If a livestream is conducted by a broadcaster of traditional media, it falls under the current Broadcasting Act 1989 and is covered by relevant Broadcasting Code Standards. However, if the livestream is conducted by an individual or group who are not a broadcaster, the conduct falls into a legislative gap.[5] For example, following the Terror Attacks, while the video recording was classified as objectionable, it was unclear whether the perpetrator had committed a criminal offence by livestreaming their attack. Because livestreaming by non-broadcasters is not covered by current legislation, Government cannot hold people or companies responsible for livestreaming objectionable content, such the Terror Attacks.

Second, **the Chief Censor must publish a written decision within five working days of**

[5] It is noted that the product of a livestream (i.e. the recording) does fall within the definition of a publication and is therefore sufficiently subject to legislative provisions.

classifying a publication, which can delay initial decisions on objectionable content. The Chief Censor has a critical role in classifying objectionable content as his/her decision makes the public aware of harm. Often, it is the Chief Censor's decision that prompts action to remove objectionable content. The FVPC Act sets out procedures for submitting publications to the Chief Censor as well as for how the Chief Censor examines a submitted publication to determine its classification status.[6] Following these procedures exactly can take time and does not suit situations where the availability of a publication is likely to be injurious to the public good and there is an urgent need to notify the public of this harm – particularly in the online sphere where media can 'go viral' quickly. In these situations, the Chief Censor may not necessarily have the time, for example, to fully justify, present all evidence, and document the decision as is required by current statutory processes; or may need to reallocate resources to do this that would otherwise be employed elsewhere, thus compromising other decisions OFLC is required to make. As a consequence, the Chief Censor may have to delay public notification in relation to a given publication. The 15 March terrorist's manifesto was a lengthy complex document and the Chief Censor had to consider delaying his classification to meet the five-day requirement.

Third, **DIA has no explicit power to request and/or enforce online content hosts to remove objectionable content from their platforms**. Neither a take-down power for objectionable content online, nor compliance measures to enforce it, were necessary in a pre-internet age. The current process for requesting the removal of objectionable content is to advise online content hosts that they may be committing an offence under New Zealand law if they do not remove the content. To date, this process is generally effective, but does not provide certainty for either Government or online content hosts and relies on goodwill and cooperation which may not necessarily be the case should similar events occur again. Following the Terror Attacks, the Government could not point to clear legislation stating that online content hosts' failure to remove the video was illegal and had no legislative mandate to compel removal of such content to prevent its spread causing harm. Companies that did comply with requests to remove content were operating without legal certainty.

Fourth, **FVPC Act penalties and the deterrent factor they play are no longer appropriate in today's digital landscape. For example, offences for non-compliance do not exist specifically for online content**. The FVPCA contains no penalties or offences in relation to hosting objectionable content online. Actors in this new landscape, such as online content hosts, do not neatly fit under traditional concepts of publishing and distributing, which apply to physical media such as books and DVDs. Additionally, financial penalties for non-compliance by large multi-national corporations are very small compared to their revenue. These penalties therefore do not serve as a deterrent.

Fifth, **the HDC Act 'safe harbour' provisions override the liability that content hosts may face under the FVPC Act**. Currently, online content hosts cannot be charged under New Zealand law for having harmful content on their websites, if they follow certain steps when someone complains. This creates potential for online content hosts being exempt from any criminal or civil liability if they break the law under the FVPC Act (which looks at more serious 'objectionable' content) but follow steps outlined in the HDC Act. This undermines enforcement efforts around objectionable content. Analysis following the Terror Attacks identified that online content hosts could simply have notified that uploader of the Terrorist's

[6] For example, Sections 3C and 3D, and 12 to 22 of the FVPC Act.

video, waited two days to take it down, and be exempted from criminal liability under the FVPC Act.

Sixth, **the FVPC Act does not provide statutory authority for blocking objectionable content online.** Government currently works cooperatively with ISPs to block child sexual exploitation material – an obvious form of objectionable online content where censorship decisions are clear-cut – via the Digital Child Exploitation Filtering System (DCEFS). Consideration of the broader spectrum of objectionable online content, all of which is illegal, deserves a clear legal framework. There is currently no statutory authority in primary legislation to support robust and transparent consideration and development of mechanisms to filter and/or block objectionable online content.

The absence of a legal framework also puts ISPs in an uncertain legal position, where currently they are making voluntary decisions to remove content from their customers' services and can be seen by their customers as censoring content. ISPs have stated they feel uncomfortable with potential legal liability for operating in this role. In the wake of the attacks, some ISPs raised concerns on this issue and continue to request greater Government support to identify what, and how, content should be blocked.

2.4 Are there any constraints on the scope for decision making?

Constraints on the scope for decision making about policy proposals are:

- Cabinet's expectation of a report back on detailed policy changes in December 2019, to allow for introduction of any legislative changes during the current electoral term;
- an expectation from civil society, ISPs and online content hosts for timely action, while at the same allowing enough time for consultation on policy changes;
- ability for DIA to take action (rather than imposing requirements for action on other agencies), restricting the scope to legislation administered by DIA;
- the cross-government Countering Violent Extremism (CVE) work programme's priorities for countering violent extremism (which are still to be endorsed by Ministers and there is a low risk of conflicting priorities); and
- expectation of the broader media regulation review to cover policy matters comprehensively, which could reduce comfort levels in endorsing changes now.

These constraints have meant that the policy proposals have been developed under time pressures and in parallel to targeted stakeholder engagement occurring, as well as significant advancements in related CVE work programmes on the part of other government agencies. The result of these constraints is that the proposals have evolved to best align with stakeholder views and complementary policy work.

2.5 What do stakeholders think?

The changes proposed are specific and technical, requiring agency and subject matter expert stakeholder consultation. There will be an opportunity for the general public to provide input as part of Select Committee processes in early 2020 (with an exposure draft provided

to technology industry groups over December 2019/January 2020).

Immediately following the Terror Attacks in March 2019, DIA interacted extensively with affected community groups, regulators and technology industry stakeholders. These interactions informed the problem definition as set out in this RIA.

DIA officials subsequently led engagement with industry, community groups, civil society and young people on problem definition and policy proposals in October and November 2019. DIA specifically posed questions about what impacts stakeholders see from violent extremist content online. The feedback covered the spectrum of content - objectionable through to offensive (but not necessarily objectionable) content that is causing harm and needs to be addressed. In engagement collateral, DIA set out the definition of objectionable and stated that violent extremist content is being interpreted within that ambit.

Stakeholders were generally supportive of the proposals, but some members of the technology industry were uncomfortable with the speed of the current process

- Most stakeholders were in support of the policy intent, and the proposals discussed during engagement. This support was contingent on the proposals adhering to three key principles:
  - Government censorship (including the processes followed) should be transparent and open to public scrutiny;
  - there must be clear, inflexible definitions of what is censored; and
  - that material should only be censored if it can be demonstrated that it causes harm.
- There was also a recurring message that while this work is useful, more needs to be done to proactively address the drivers of violent extremist behaviour and provide protection for groups who are targeted by it.
- Members of technology industry groups asked for additional detail (as the proposals were intentionally presented at a high level) on the proposals, to understand how they would impact their businesses and to avoid unintended consequences. They further queried whether timeframes could allow an opportunity for general public consultation and for more detail to be provided.

In addition to the stakeholders noted above, DIA consulted with OFLC, as well as the following agencies on proposals in this paper: Ministry for Culture and Heritage (MCH); Ministry of Foreign Affairs and Trade (MFAT); Ministry of Business, Innovation and Employment (MBIE); Ministry of Justice (MoJ). Their feedback has informed development of options and analysis and is summarised below.

- **MBIE:** Suggested scoping the proposed changes/options to deal with a very specific problem (preventing a repeat of the Terror Attacks), and that future proofing changes through broader principles application, beyond creating the mechanical settings, may assist with this.
- **MCH**: Suggested considering the impact of the proposals on the upcoming broader media regulation review and whether they would be in or out of scope for that review.
- **MFAT**: Suggested providing detail on the practicalities of the legislation and the nature of the issues at hand; differentiating between problems arising from the Terror Attacks vs ageing legislation; defining terms upfront; ensuring non-regulatory measures are included; and describing how a take-down notice would work (e.g. whether it would apply to just one URL and/or to mainstream media as well as social media companies).

- **MoJ**: Suggested considering giving more weighting to United States enforcement (where many large online content hosts are based); considering using a civil pecuniary penalty instead of a new criminal offence for large corporate non-compliance; and advised there was no need to increase the penalty level for the current offence.
- **OFLC:** Suggested providing detail on the operational funding agreed in September and joint monitoring arrangements with DIA; including the specific section of the FVPC Act that relates to violent extremism; clarifying the act of livestreaming rather than the distribution of a recorded livestream is not currently an offence; clarifying that no blocking mechanism can be 100% effective; and providing further detail on the differing views within civil society groups.

As a result of feedback, DIA has:

- extended the engagement timeframe and will supply technology industry groups with an exposure draft of the legislation to comment on over December 2019- January 2020;
- clarified that the proposals apply to objectionable content; and
- made clear in proposals the intention for transparent and routine reporting on take-down notices, as well as clarifying accountability and transparency measures (for example, consultation requirements, decision-making processes and appeal and review pathways) in relation to any potential mechanisms for filtering and/or blocking objectionable content.

Section 3: Options identification

3.1 What options are available to address the problem?

We have identified four options to address the problem. The status quo is outlined as Option 1. The proposed suite of regulatory changes to improve our response to objectionable content online is outlined below as Option 2. A third option is the suite of legislative options outlined under Option 2 but applied only to violent extremist content (and excluding other objectionable content). A fourth option is a non-regulatory partnership approach based on a Memorandum of Understanding (MoU).

There are few dependencies between options. Option 4 (MoU) could be progressed alongside any other option. Options 2 and 3 require regulatory change. Option 1 (status quo) can be progressed in addition to undertaking regulatory changes in future (Options 2 or 3). Option 3 would likely prompt definitional issues that could lead to decisions to adopt changes more in line with Option 2.

We have included our analysis of several options ruled out as not viable at this time, for completeness. To mitigate risk, regulatory changes will be underpinned by ongoing and continued engagement with ISPs and online content hosts.

Option 1: Status quo

Under this option, no regulatory interventions would occur immediately. While DIA and OFLC would still benefit from the recent operational uplift in capacity to identify and assess objectionable content online that promoted and supported violent extremism, actions to counter such content would remain constrained in the absence of regulatory changes to support a better resourced operational response. This would result in New Zealanders remaining at risk of harm from exposure to objectionable content online.

Regulatory options to address these issues would be deferred for consideration as part of the broader media regulation review, meaning the earliest any changes would be implemented is 2022.

Actions would be limited to non-regulatory approaches currently being pursued under the Christchurch Call, involving active dialogue between online content hosts, ISPs and governments to take action to counter objectionable content online. These actions rest on a non-binding, goodwill-based joint commitment by Government and the technology industry, and their effectiveness or otherwise will be determined by the ongoing cooperation of ISPs and online content hosts, with few levers for Government to draw on should such cooperation cease.

Option 2: Suite of improved regulatory levers

Under this option, the Government would make a suite of six regulatory changes to the FVPC Act. These changes would enhance the ability of government and non-government partners to respond to deter and respond to objectionable content online. The options forming the suite are not intended to be considered as standalone, due to the technical and interconnected nature of the proposals. Progressing the proposed suite of changes in the near term is not anticipated to delay the planned review of the media content regulatory system, expected to begin in 2020.

Create a new offence for livestreaming objectionable content

This change seeks to add a new offence under the FVPC Act to criminalise 'livestreaming' (i.e. broadcasting over the internet in real time) of content that, if it fell under the definition of publication in the Act, would be considered objectionable. The proposed offence will specify that people or companies livestreaming such content will be subject to criminal penalties and provide legal clarity. Immediately following the Terrorist Attacks, it was unclear whether the act of livestreaming the attack footage was a criminal act or not.[7] This proposed offence would not apply to online content hosts, only the individual/group initiating the livestream of the content.

Grant the Chief Censor authority to make interim classification decisions

This change seeks to amend relevant sections of the FVPC Act so that in circumstances where the availability of the publication is likely to be injurious to the public good, and an urgent need exists to notify the public of this harm, the Chief Censor would have the authority to make an interim decision on classification without activating the five-working day limit for a full written decision. The interim decision would have the same effect as a written decision. The interim decision would be in place for a maximum of 20 working days before the Chief Censor would be required to issue a final decision.

This option would empower the Chief Censor to indicate a likely classification status, and alert the public and enforcement agencies accordingly, when an urgent need exists to notify the public. This means the public enforcement agencies and online content hosts would be better prepared sooner for how to treat such content. This would minimise the rate at which

[7] Note the product of livestreaming (i.e. the recording) can be considered a publication and therefore subject to classification as objectionable.

violent extremist and terrorist material spreads online and reduce instances of people being negatively affected by harmful content.

A final decision would still be required within 20 working days, which would limit the uncertainty of an interim decision versus a final decision.

The amendment would need to protect the Chief Censor, ISPs and online content hosts, from criminal or civil liability from actions relating to such an interim decision. This is so that ISPs and online content hosts can proactively remove/block publications in advance of a final decision without fear of repercussions from making the "wrong move" (e.g. being subject to a legal suit), and to avoid an employee or website moderator being seen as committing an offence when dealing with violent extremist and terrorist content when tasked with assessing, referring to or removing reported violent extremist and terrorist material from a website.

Grant DIA inspectors take-down powers for objectionable content online

This change seeks to amend the FVPC Act to provide inspectors of publications (either from DIA or the New Zealand Police) with the power to issue take-down notice where they identify, either from their own interpretation of the Act, or based on the Chief Censor's decision, objectionable content online. This power would be in addition to the inspectors' current practice of requesting voluntary take down of content in individual cases. Where an inspector of publications issues a take-down notice based on their own interpretation of the Act, they will notify the Chief Censor who can choose to classify the publication to confirm its status. The form of takedown notice will be prescribed in regulations.

To ensure accountability in the use of this function, the Secretary of the Department of Internal Affairs (the Secretary) will be required to publish a list of all take down notices complied with in a way that is accessible to the public. The Secretary will also report on take down notices issued as part of its annual reporting requirements.

The inspectors' take-down powers would be in line with current powers of removal of objectionable content[8] under the FVPC Act. Inspectors would issue a take-down notice to an online content host or online platform, directing removal of a specific link, so that the relevant material could no longer be viewed by people accessing it from New Zealand, thereby reducing harm from this content. This would require online content hosts to either remove content, or to make the content unavailable to people attempting to access the content with a New Zealand IP address. We acknowledge that this mechanism could be circumvented using several technical methods, including a virtual private network.

An online content host that did not comply with a notice to take down content within 24 hours of receiving notice (or alternative time set out in legislation) would be committing a breach, as described in the FVPC Act (see later section).This change would be consistent with the German *Network Enforcement Act (NetzDG)*, which requires large technology companies to remove "manifestly unlawful" content within 24 hours of receiving a complaint. Research

[8] Section 108, the Films, Videos, and Publications Classification Act 1993.

suggests that NetzDG's most important effect was to ensure swifter and more consistent removal of content within Germany, under the companies' community guidelines.[9]

In order to impose a penalty, a specific timeframe is necessary. A 24-hour timeframe strikes a balance between a swift timeframe for action and sufficient time for the online content hosts to comply with a notice.

The intent of this change is to bring online content hosts in line with the expectations of businesses operating in New Zealand with respect to dealing with content classified as objectionable. Liability for this breach would necessarily be limited to content hosts and online platforms that are liable under New Zealand law. The amendment would need to include protections for employees of ISPs and online content hosts who are tasked with investigating and/or taking down objectionable content online from criminal or civil liability. This is to avoid an employee or website moderator being seen as committing a breach when dealing with objectionable content online when tasked with assessing, referring to or removing reported objectionable material from a website.

Apply penalties to online content hosts for non-compliance

This change seeks to amend Part 8 of the FVPC Act to introduce the application of a civil pecuniary penalty for online content hosts where they do not comply with a take-down notice relating to objectionable content online.

This would mean that, in the event of non-compliance with a take-down notice, a civil pecuniary penalty could be applied. The intent of this change is to bring online content hosts in line with the expectations of businesses operating in New Zealand.

It will be a defence for an online content host to demonstrate reasonable justification where they are unable to meet the 24-hour deadline.

The amendments need to apply to both resident and overseas based online content hosts, if their content is available in New Zealand. In practice, neither civil nor criminal penalties may be enforceable if the online content host concerned does not fall within New Zealand's jurisdiction. We note there is a Reciprocal Enforcements of Judgements legislation that applies to some countries such as Australia. However, take-down notices and penalties for non-compliance create greater transparency for online content hosts who, in general, are reluctant to incur any reputational risk associated with any infringement of the regulations of the countries they operate in.

Override the HDC Act 'safe harbour' provisions when the FVPC Act applies

This change seeks to amend the FVPC Act to expressly override section 24 of the HDC Act, as is provided for in section 25(4)(d) of that Act. No amendments would need to be made to the HDC Act.

This would mean that the Secretary could enforce any offence in the FVPC Act and not be affected or limited by the HDC's 'safe harbour' provisions for content hosts. It would ensure that online content hosts could be prosecuted for possessing or distributing objectionable or

[9] "An Analysis of Germany's NetzDG Law" Heidi Tworek, University of British Columbia, Paddy Leerssen, Institute for Information Law, University of Amsterdam, April 15 2019, https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf

restricted content in relation to Violent Extremist Online Content.

This change is closely tied to the liability of overseas-based online content hosts. This change on its own would have a limited impact if these companies are not liable under New Zealand law.

Enable establishment of possible mechanisms for filtering and blocking objectionable content online

This change seeks to amend the FVPC Act to enable, subject to further scoping and consultation, establishment of mechanisms such as a mandatory web filter at ISP level to block objectionable content online. The filter or blocking mechanism would help limit exposure to objectionable content online, thereby limiting harm to New Zealanders.

This amendment would give Government explicit statutory authority to implement such mechanisms through regulations. The detail of the mechanism would be set out in future regulations. Due to the complexity of filtering and blocking mechanisms, the Minister of Internal Affairs and/or the Secretary will be required to consult with ISPs and online content hosts to design such a mechanism and must consider with relevant agencies the impacts on consumer access to the internet to non-objectionable material, preserving freedom of expression rights and other relevant constitutional or legal issues, and compliance costs in implementing filtering or blocking obligations.

The amendment to the FVPC Act (the principal Act) would:

- articulate the purpose of such a mechanism/s as being to limit exposure of people in New Zealand to objectionable content online;
- clarify that, filtering and/or blocking will apply to objectionable content online;
- provide for the power for DIA to operate such a mechanism/s for filtering and blocking online content as a regulator;
- provide for the requirement to consult on the design of the mechanism; and
- provide for regulations to be developed that establish the detail of any mechanism.

The amended Act would require that regulations would need to describe:

- the scope and requirements for filtering websites;
- data handling and privacy provisions;
- decision-making and governance arrangements, including any requirements for Ministerial and officials' involvement, as well as independent experts;
- the duties/obligations of ISPs and online content hosts to comply with statutory requirements;
- the review process and right of appeal should an ISP, online content host or other individual or entity dispute a decision;
- appropriate protections from civil or criminal liability for officers of DIA, law enforcement or reporting entities or others who would interact with the content while complying with legislative provisions; and
- the ability for the Secretary to issue codes of practice in relation to the mechanism.

This amendment would not be necessary if mechanisms were voluntary (rather than compulsory) and narrowly scoped, as is the case for the existing DCEFS. This is because content in scope for the DCEFS is relatively easy to make a decision on, whereas decisions

on objectionable content can be more discretionary and complex; 'objectionable' covers a wider range of illegal content than only child sexual exploitation material. We would need to consider freedom of expression implications, and other unintended consequences of filtering a broader range of online content e.g. removal of content that has cultural and or evidential significance or impacts on shared internet infrastructure that may be affected by broad blocking/filtering. Because of this, appropriate ISP and online content host input will be a crucial element into the design of such a mechanism.

Option 3: Narrow focus on legislative change to focus on violent extremist and terrorist content online only

Under this option, the proposals in Option 2 would apply to violent extremist and terrorist content online only, rather than all objectionable content online. Under this option, the new regulatory functions would focus on mitigating harm from content such as the Terror Attacks video and manifesto, at the exclusion of other types of harmful objectionable content, such as child sexual exploitation material.

Option 4: MoU-based partnership approach (non-regulatory)

Under this option, there would be no changes to legislation. The Government would instead seek to augment and expand current and emerging industry partnerships established under the Christchurch Call via a formal MoU with ISPs and online content hosts.

The obligations set out under the MoU would be similar to those set out under the proposed suite of legislative options but would not be set in law, and therefore likely include:

- a clear outline of the responsibilities of Government and of ISPs and online content hosts in relation to objectionable content online;
- an agreement that online content hosts will remove content from their platforms once they receive a request from Government to do so;
- an agreement to promote positive behaviour online such as an educative function through re-direction of users from objectionable content to sites providing people with support; and
- detail of any agreed voluntary filtering arrangement (noting this would likely take time to negotiate).

As for Option 1, unless ISPs and online content hosts agreed to a binding MoU, the effectiveness of commitments would continue to rest on the goodwill of ISPs and online content hosts, with no regulatory levers for Government to employ should cooperation cease. DIA and OFLC would still benefit from the recent operational uplift in capacity to identify and assess objectionable content online, but actions to counter such content would remain constrained in the absence of policy changes to support a better resourced operational response. This would result in people in New Zealand remaining at risk of harm from exposure to objectionable content online.

This option could be progressed alongside either of the regulatory approaches described in Option 2 and Option 3.

3.2 What criteria, in addition to monetary costs and benefits, have been used to assess the likely impacts of the options under consideration?

Assessment criteria

1. **Effectiveness of intervention**: enables both government and private actors to swiftly assess and classify objectionable content online and to rapidly remove such content from online platforms.
2. **Consistency with the Government's broader CVE agenda**: is consistent with the broader counter-terrorism work programme, democratic and social norms, our international commitments and other applicable standards.
3. **Wellbeing and security**: enables people in New Zealand to maintain their wellbeing and ensure their safety and security.
4. **Ease of implementation**: minimises implementation barriers or complexities; delivers a solution that can be monitored, enforced, and delivered in a timely and effective manner.
5. **Efficiency**: is proportionate to the risk, easier to understand and apply; benefits of the option expected to exceed the costs.
6. **Stakeholder (industry/civil society) relationships and certainty**: supports positive and enabling partnerships and provides regulatory certainty for the industry, civil society and tech bodies.

Trade-offs

The assessment criteria are largely complementary. At times, there may need to be trade-offs, for example between providing DIA and OLFC with greater regulatory powers and increased industry compliance costs.

Increasing regulatory requirements will be balanced against the Christchurch Call's commitment to a partnership approach.

3.3 What other options have been ruled out of scope, or not considered, and why?

As discussed in section 2.4 above, there are several interdependencies related to this work programme. Proposals that fit under separate workstreams were ruled out of scope. The following options were also identified but not considered viable.

a. Extend the Chief Censor's mandatory classification reporting time in relation to objectionable content online

Description: The FVPC Act could be amended to extend time the Chief Censor has to issue a written decision from five working days to a maximum of 20 working days.
Assessment: This option is not considered viable because it is unlikely to reduce the harm resulting from the online dissemination of objectionable content online if there is no obligation for online content hosts to remove the content until a written decision has been published. Further, an extended reporting timeframe may create more public uncertainty as to the status of the publication and the legality of viewing/possessing a copy.

b. Transfer decision-making powers on certain objectionable content to another agency

Description: The FVPC Act could be amended to transfer decision-making powers on objectionable content relating to violent extremism online from the OFLC to another agency (e.g. DIA or Police) for faster processing and classification. This would likely require setting

up a new unit within an existing agency.
Assessment: This option is not considered viable because it would involve significant resourcing to upskill the relevant agency and there is considerable risk of fragmentation and confusion of responsibilities between two different units. There is the danger of Government being seen to censor unpopular views; OFLC brings a level of independence from Government. The OFLC was established as an independent entity to balance the fundamental rights and freedoms of New Zealanders with public safety and the wider public good.

c. *Amend the FVPC Act to grant DIA inspectors take-down powers based on a Court order*

Description: DIA Inspectors would require the Court's approval to request the take-down of content. The Court would consider the evidence provided as to whether the website in question met all necessary requirements for taking down content (that it is objectionable and is displayed to New Zealanders).
Assessment: This option is not considered viable because it would have a slower response time and requires a much higher bar than the status quo. The approach is likely to be more resource and cost intensive, with the potential for reduced effectiveness, depending on online content hosts' liability under New Zealand law. It would also increase the burden on courts through additional workload, which could have a flow on effect of affecting timely resolution of other high priority legal cases.

d. *Amend section 25(4) of the HDC Act to provide a safe harbour for offences under the FVPC Act*

Description: Section 25(4) of the HDC Act would be amended to include offences under the FVPC Act in the list of Acts that are not affected by sections 23 and 24 of the HDC Act.
Assessment: This option is not considered preferable because the same outcome can be achieved by amending the FVPC Act. If changes are already being made to that Act it makes sense to amend only one rather than two pieces of legislation. It is also preferable for provisions relating to objectionable content to sit in the primary Act that deals with this matter.

e. *Establish an independent regulatory body*

Description: A new independent regulator would be set up along the lines of the New Zealand Media Council and Broadcasting Standards Authority. This would replace the current regime in which social media companies are largely left to self-regulate how they monitor and remove harmful content.
Assessment: This option is considered not viable because it would raise a range of complex issues relating to consistency and relationships across different media platforms that would be better considered in the context of the broader media regulation review.

Section 4: Impact Analysis

Marginal impact: How do the options at section 3.1 compare with the counterfactual, under each of the criteria set out in section 3.2?

Note: We recognise that Options 2 and 3 comprise six discrete legislative proposals. In developing this RIA, we assessed each proposal individually against the criteria. We have chosen to present analysis against the criteria in summary given the level of duplication across all amendments, and to underpin the nature of the option as an interconnected suite of amendments, as opposed to amendments intended for introduction individually or as standalone solutions.

Option	1 (no action, status quo)	2 (Suite of legislative changes)	3 (Legislative changes for violent extremist and terrorist content only)	4 (MoU-based partnership approach)
Effectiveness of intervention	**0**	**++** Comprehensive, tightly-scoped interim response. Enterprise level approach to objectionable material to enable rapid response	**+** Enhanced Government interventions limited to VE forms of objectionable content	**+** Government still limited in the ability to classify, remove and block content. Dependant on ongoing goodwill of non-government partners.
Consistency with broader CVE agenda	**0**	**++** Supports Christchurch Call and broader CVE government programme while introducing immediate tangible enablers	**++** Supports Christchurch Call and broader CVE government programme while introducing immediate tangible enablers	**+** Actively builds on and expands Government/industry partnership established under the Christchurch Call
Wellbeing and security	**0**	**++** Combined changes will supply both tangible tools and sense/premise of increased security and wellbeing	**+** Provides stronger premise for wellbeing and security, but may be viewed as inadequate as interventions limited to specific types of harmful content i.e. VE	**+** Provides stronger premise for wellbeing and security, but may be viewed as inadequate by civil society groups expecting stronger interventions from Government
Ease of implementation	**0**	**+** Some of the proposals may be more challenging to implement than others. External factors may impact on success of implementation	**+** Operationally difficult to implement, creates anomalies within the FVPC Act, could be confusing for ISPs and OCHs	**+** Challenging to address non-compliance with take-down notices
Efficiency	**0**	**++** The suite presents a comprehensive response with societal benefits, at a proportionate cost	**+** Comprehensive response to terror attacks, but not proactive for future situations of this magnitude	**0** Low cost, but limitations remain
Stakeholder (industry/civil society) relationships and certainty	**0**	**++** Provides comprehensive legal certainty and guidance	**+** Provides legal certainty around particular types of content (VE), but could also create uncertainty and confusion around other types of objectionable content	**+** Provides greater guidance, but not legal certainty, which was requested by industry following the Terror Attacks.
Overall assessment	**0**	**++** Bolder, comprehensive course of action that will future proof against potential delays in broader media regulation review	**+** Targeted response to violent extremist content, but inconsistent with DIA's approach to regulating objectionable content. May create confusion and a 'hierarchy of harms' – i.e. a public perception that a	**+** Partnership approach likely to be supported by industry, but fall short of civil society expectations and not improve the Government's ability to respond to a situation similar to the Terror Attacks

Option	1 (no action, status quo)	2 (Suite of legislative changes)	3 (Legislative changes for violent extremist and terrorist content only)	4 (MoU-based partnership approach)
			focus on VE harms is more important than a focus on child sexual exploitation material	

Key: **++** much better than doing nothing/the status quo | **+** better than doing nothing/the status quo | **0** about the same as doing nothing/the status quo | **-** worse than doing nothing/the status quo | **- -** much worse than doing nothing/the status quo

Section 5: Conclusions

5.1 What option, or combination of options, is likely to best to address the problem, meet the policy objectives and deliver the highest net benefits?

No action or deferring changes until the broader media regulation review (i.e. the status quo of Option 1), are not acceptable options because they would mean government and industry remain ill equipped to respond to objectionable content online, including violent extremist and terrorist content until that review is completed. The status quo also fails to meet community and Government expectations of rapid and active intervention to mitigate against future harm arising from objectionable content online, as was expressed clearly through engagement.

Option 2 (suite of regulatory changes to the FVPC Act) is the preferred option because it will enable a faster response to prevent harm from objectionable content online through:

- new and improved processes and powers to support effective and timely censorship decision-making and enforcement;
- greater clarity and guidance for technology partners on expectations and responsibilities in responding to objectionable content online; and
- scope for investigating and establishing future mechanisms to filter and block objectionable content online, subject to further policy development and consultation.

This suite of changes is consistent with the broader government CVE agenda, will increase wellbeing and security of people in New Zealand, and has the potential to improve industry relationships and certainty. Implementing changes and ensuring they are efficient is not likely to be any more challenging than the status quo, given the current context post the Terror Attacks.

Option 3 (suite of regulatory changes to the FVPC Act applying to violent extremist and terrorist content only) would provide greater legal clarity and enforcement functions relating to violent extremist content but would not support consistency in the application of new functions. For example, operational staff would be able to issue take-down notices to platforms hosting violent extremist content, but not to platforms hosting child sexual exploitation material. Furthermore, a narrow focus on violent extremism in relation to the new functions would also send an incorrect public message that violent extremist/terrorist harms are of a higher priority than harms arising from other objectionable material, such as CSEM. This perception may also lead to public confusion about what DIA is able or willing to focus on.

Option 4 (an MoU-based partnership approach) would have the greatest positive impact on Government relationships with ISPs and online content hosts. However, this option could be seen as:

- not sufficiently action-oriented to equip agencies with tools required to respond to objectionable content online when required; and
- not addressing legislative ambiguity for industry partners.

We have canvassed these options with relevant teams in DIA, including Digital Safety and Legal, who support the policy changes. We have also received positive and constructive feedback on initial high-level outlines of the options for change to address limitations in domestic policy settings from OFLC and other agencies involved in the wider Government counter-terrorism and national security efforts, including DPMC, MFAT, MoJ and MCH.

We have taken on board feedback received from industry and community groups during engagement in October and November 2019. All this feedback has informed development of these options into the suite of reforms forming the preferred option. The options forming the suite are not intended to be considered as standalone, due to the technical and interconnected nature of the proposals.

5.2 Summary table of indicative costs and benefits of the preferred approach

Affected parties *(identify)*	**Comment** *nature of cost or benefit (eg ongoing, one-off), evidence and assumption (eg compliance rates), risks*	**Impact** *$m present value, for monetised impacts; high, medium or low for non-monetised impacts*	**Evidence certainty** *(High, medium or low)*
Additional costs of proposed approach, compared to taking no action			
Regulated parties	Disruption from adjustments to new regulatory settings; ongoing 'hosting liability' costs (e.g. for moderators)	Low-medium	Low
Regulators*	DIA and OFLC establishment costs (met through funding approved on 30 September)	Existing FTE	Medium
Wider government	Increased engagement with technology partners and other stakeholders	Medium	Medium
Civil society	Ethnic communities and religions – risk of backlash (cost)	Medium	Medium
Total non-monetised cost	**Implementation, engagement, backlash**	**Medium**	**Medium**
Expected benefits of proposed approach, compared to taking no action			
Regulated parties	Increased certainty and reduced wasted effort through greater clarity of processes and expectations	High	Medium
Regulators	Ability to more effectively respond to objectionable content and therefore reduce / mitigate harm	High	Medium
Wider government	OFLC (independent Crown entity) – increased certainty and clarity of how to respond	Medium	High
Civil society	Better protection for ethnic communities and religious groups online; increased wellbeing and security through more effective and enabling regulatory settings	High	Medium- due to pluralistic nature of civil society
New Zealand public	Reduced exposure to objectionable content online	High	Medium - due to differing levels of exposure to harmful

			content within the general public
Total non-monetised benefits	**Reduced risk of harm to civil society from exposure to objectionable content online.**	**Medium**	**Medium**

* Any costs associated with the implementation of a filtering mechanism will be considered in a subsequent Regulatory Impact Assessment.

5.3 What other impacts is this approach likely to have?

Due to the complex and outdated nature of the current media content regulation regime, changes to the FVPC Act may have small unintended impacts on other pieces of legislation. This will be thoroughly tested through agency consultation and the broader media regulation review.

Potential risks and uncertainties

Under the status quo, there is a significant risk of harm for New Zealanders exposed to objectionable content online, whether originating from within or outside the country.

It is difficult to predict the extent to which these changes would be effective should another event of a similar magnitude to the Terror Attacks occur but involving a different category of online objectionable content. However, it does provide more tools and more certainty for both government regulators and ISPs and online content hosts.

There are several risks associated with the proposed changes, as outlined below. We consider we can mitigate these risks appropriately.

- **Risk that new decision-making powers around classification, interception and removal of online content, or new offences, are applied in situations other than intended or envisaged.** We can mitigate this risk by ensuring that the preferred suite of options is specific to objectionable content, as verified by established reporting requirements. Furthermore, these powers would make clear that news outlets would not be liable for reporting from the site of the tragedy, provided that the footage is not graphic and in the public interest (this approach is consistent with recent Broadcasting Standards Authority decisions).[10]
- **Risk that ISPs and online content hosts react negatively to more restrictive regulation of online content**. We can mitigate this risk by limiting the scope of the proposed regulatory changes to objectionable content. We can further mitigate this risk by engaging collaboratively with key industry partners (i.e. online content hosts and ISPs) about the form and implementation of the regulatory changes in accordance with

[10] The BSA upheld a complaint regarding the Sky News broadcast of 'disturbing, violent content' from the mosque shooter's video, but not complaints regarding TVNZ's use of brief, non-graphic clips of the video or footage of the victims.
See https://bsa.govt.nz/decisions/all-decisions/uj-and-sky-network-television-ltd/.
See https://bsa.govt.nz/decisions/all-decisions/nt-and-television-new-zealand-ltd/ and https://bsa.govt.nz/decisions/all-decisions/grant-and-phillips-and-television-new-zealand-ltd-2019-013/.

the Christchurch Call and international best practice. We would specify that new penalties or offences apply only to non-compliant online parties. Major ISPs and online content hosts have already demonstrated goodwill and their support of reasonable, implementable changes. The risk of a disproportionate reaction, such as their withdrawing from either the current voluntary mechanisms underpinning the Christchurch Call, or the New Zealand market as a whole, is low. For example:

- Amazon, Facebook, Google, Twitter and Microsoft have issued a joint statement outlining the nine steps they will take to implement the Christchurch Call.[11]
- On 17 September 2019 Facebook announced that anyone in New Zealand searching its site for extremist content will in future be directed to websites helping people to leave hate groups.[12]
- On 24 September 2019 the Global Internet Forum to Counter Terrorism (GIFCT), set up by Facebook, Microsoft, Twitter and YouTube, announced a new broader vision to prevent terrorists and violent extremists from exploiting members' platforms.[13]

However, we acknowledge that these companies are not representative of all ISPs and online content hosts, who may have ongoing collaboration with the New Zealand Government on different service areas, for example cloud storage. These proposals could have the unintended consequence of negatively impacting these partnerships.

- **Risk that ISPs and online content hosts become more risk averse and restrict material that is not objectionable content**. This could happen if an unreasonably restrictive regulatory approach for access to online content were established, such as a wide-ranging web filter without strong governance and appeals processes. This would diminish the utility New Zealanders get from online content platforms without necessarily reducing harm associated with exposure to objectionable content. We can mitigate this risk by consulting with industry partners and experts to develop the detail of regulatory changes, and maintaining open and productive communication channels with industry to be able to work through issues if or when they arise.
- **Risk that take-down powers are ineffective or have adverse impacts**. Take-down notices may be unnecessary (e.g. larger companies who are likely to observe the Chief Censor's classification decisions and remove content as requested or proactively), or ineffective (e.g. smaller companies who are unlikely to comply with decisions on violent extremist and terrorist content, who may ignore notices and actively seek to pervert classification decisions by harnessing violent extremist sympathisers and incite further attacks). We also recognise that notices to take down content apply only to information published on websites and are not effective in stopping peer-to-peer (P2P) content sharing, which is a growing trend. We can partially mitigate this risk by establishing protocols for when take-down notices would be issued (e.g. it may only be necessary in some situations, which would reduce administrative burden), and by assessing the risk of requesting the take-down of content in each circumstance. As large online content hosts are often reluctant to run the reputational risk of openly flouting the laws of the countries in which they operate, and the obligations on them are quite clear, there is

[11] https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2019/05/Christchurch-Call-and-Nine-Steps.pdf
[12] https://newsroom.fb.com/news/2019/09/combating-hate-and-extremism/
[13] https://www.beehive.govt.nz/release/significant-progress-made-eliminating-terrorist-content-online

little reason to assume that non-compliance will be their default position. Furthermore, we would establish regular and transparent reporting on the number of take-down notices and the outcome (e.g. whether the objectionable content was removed and how long it took to remove the content).

- **Risk that offshore-domiciled companies choose to escape liability introduced by new or clarified penalties and offences by not answering the charges**. Online content hosts may maintain that they are immune from any liability for the online content of others, particularly those based entirely outside of New Zealand who are unable to stop people in New Zealand from accessing their content. This is a reflection of the reality that the borderless internet does not fit neatly into geographically defined jurisdictions. Furthermore, many online content hosts are based overseas without a New Zealand footprint, which means they cannot be found liable under New Zealand law. We can partially mitigate this risk by imposing civil pecuniary penalties that can be enforced either in New Zealand, or by other countries where mutual agreements support this (e.g. the Reciprocal Enforcement of Orders Act provides for the enforcement of New Zealand civil orders in Australia).[14] The fact of setting a standard, even if difficult to enforce, will signal that the New Zealand Government is serious about removing objectionable content online and expects online content hosts to share this responsibility.
- **Risk that some civil society groups view the changes as unduly limiting freedom of expression.** While some civil society groups will agree with the changes and potentially even think Government could regulate further, some will raise concerns of infringing freedom of expression. We can mitigate this risk by clearly communicating the Government's position that objectionable (i.e. illegal) content online should be eliminated: this is the most appropriate response for the safety and security of all New Zealanders. We can mitigate this risk further by requiring transparent reporting on the use of new powers to ensure that Government/regulators are accountable for use of new powers.
- **Risk that the changes are viewed by stakeholders as premature and more appropriately considered as part of the broader media content regulation review, leading to erosion of support and trust for that review.** We can mitigate this risk by articulating that a delay would be inconsistent with the Government's commitments under the Christchurch Call to address legislative limitations to counter violent extremist content online. Furthermore, the effectiveness of the current proposals could be considered as part of the broader media content regulation review. We can reassure stakeholders of intended engagement as part of that review.
- **Risk that the proposed changes are inconsistent with the partnership approach agreed through the Christchurch Call.** We can mitigate this risk by ensuring continued open communication with stakeholders, including other government agencies, ISPs and online content hosts. The CVE Ministerial Group will ensure the proposals are consistent with the Christchurch Call and other related government initiatives. This risk has been realised: some industry partners have expressed disappointment with the short and limited consultation on proposals.
- **Risk that mechanisms for blocking objectionable content are seen variously as ineffective, easy to circumvent or infringing on freedoms of expression.** The

[14] See the Trans-Tasman Legal Proceedings Act 2010 http://www.legislation.govt.nz/act/public/2010/0108/latest/whole.html#DLM2576223

internet is borderless and there are mechanisms for people to circumvent controls (e.g. under German regulations information is hidden to Facebook users who say they are in Germany, but is not actually removed). We acknowledge that the proposed reforms will not prevent individuals and groups from actively seeking out objectionable content on the internet. However, we can mitigate the risk of people finding it accidentally by working in partnership with and relying on the technical expertise of industry partners, as well as government officials and experts, to investigate the most appropriate and effective mechanisms for filtering and removing such content (subject to further consultation). As with our intended approach for take-down notices, it would be important as part of regulatory design to establish regular and transparent reporting arrangements on the material filtered or blocked by a mechanism.

There is no example internationally of a government that has successfully balanced freedom of speech and political expression, with the need to minimise harm from objectionable content.[15] The regulatory and legislative regimes enacted by other governments have had mixed effects and reactions. It is broadly acknowledged that the most effective approach to defining and agreeing the social responsibilities of technology industry bodies (including social media companies) is to create a private-public dialogue between Government and internet intermediaries themselves. This approach is consistent with commitments under the Christchurch Call.

A successful non-regulatory approach to continued engagement will comprise:

- *Clear communications and expectations*: Government would ensure overseas-based online content hosts know their liability under New Zealand law if they publish or distribute objectionable content. Government would clearly set out requirements under relevant legislation for entities operating from a location outside New Zealand, what the Governments expects (even if not legally able to enforce given extra-jurisdictional challenges), and development of a framework for cooperation.
- *Opportunities for input to future regulatory design, towards a co-regulatory approach*: Government would seek to ensure opportunities for continued meaningful and informative collaboration between government and non-government partners relating to online media content, including but not limited to objectionable content. This collaboration would seek to balance freedom of expression and ensuring that human rights remain protected from harm (e.g. terrorism and violent extremism). For example, there are initiatives worldwide to develop a binding 'duty of care' for internet intermediaries which could be explored further beyond the recommended option.

[15] **Germany**: enacted the Network Enforcement Act 2017, which was criticised by international organisations for being both too far reaching and having limited effect on the spread of hate speech. **Australia:** legislated a new offence after the 15 march attacks relating to the expeditious removal of 'abhorrent violent material'. The law has been subject to intense criticism, arising from the vague definitions of "expeditiously" and "abhorrent violent material", and ambiguity of enforcement. **European Union:** approved legislation to introduce a requirement that platforms remove terrorist content within 12 hours of receiving a removal order. There are concerns that it does not strike the right balance between free speech and preventing terrorism content. **United Kingdom:** recently published the Online Harms White Paper, which sets out the government's plans for a package of online safety measures, that would make companies more responsible for their users' safety online, especially children and other vulnerable groups.

5.4 Is the preferred option compatible with the Government's 'Expectations for the design of regulatory systems'?

The preferred option of a suite of reforms is expected to deliver net benefits to New Zealanders in the immediate and longer term and is compatible with the Government's 'Expectations for the design of regulatory systems'.

This is because the preferred option:

- has specific objectives (i.e. addressing the issues articulated in section 2.3) and is a proportionate response to the Terror Attacks, focussing on immediate actionable changes to processes as well as improving domestic regulatory settings and strengthening productive partnerships with industry bodies;
- seeks to have a minimal impact on market competition and individual responsibility, balanced against the rights of all New Zealanders to safety and security (discussed at section 5.3);
- provides clear expectations and guidance for key parties involved in responding to objectionable content online, i.e. DIA, OFLC, online content hosts, ISPs and civil society (per the option description at section 3.1 Option 2);
- seeks to reduce unintended gaps in the media content regulatory framework through updating the FVPC Act (per the option description at section 3.1 Option 2);
- establishes statutory authority to create a regulatory mechanism for filtering or blocking objectionable content without actually creating this mechanism now, sets out expectations for further consultation on design and implementation of such a mechanism, and has scope to evolve as needed (per option description at section 31. Option 2);
- provides flexibility for regulators to meet current societal expectations through new powers, and envisages continued partnership between regulators and regulated parties on design and implementation of a filtering or blocking mechanism (refer sections 3.1 Option 2, 5.1 and 2.5);
- conforms to established legal and constitutional principles and supports compliance with New Zealand's international and Treaty of Waitangi obligations (per reference in Section 3.1 Option 2 to the need for careful navigation of constitutional and legal issues from filtering/blocking mechanisms, and consideration of such issues as implementation risks at section 6.2).

Section 6: Implementation and operation

6.1 How will the new arrangements work in practice?

The FVPC Act needs to be amended to give effect to the proposed legislative changes. It is intended that the Act would be amended by September 2020, following the Bill being passed into legislation by August 2020. This is however largely dependent on the House programme and priorities.

We recognise that the effectiveness of the proposed suite of changes rests on robustly planned implementation. DIA will lead implementation of the proposals, in consultation with industry, and relevant government agencies. DIA will work closely with relevant partners at an operational level to define new/adjusted processes for classification decision-making

timeframes and take-down powers, and how best to articulate a chain of escalation. It is envisaged that an appropriate sequence could be:

- initially, DIA officials make an informal take-down request (e.g. via email), as is current practice;
- the next step, if necessary, would be DIA officials issuing a formal take-down notice, potentially but not necessarily supported by an OFLC classification decision;
- the step after that, if necessary, would be DIA giving notice to the online content host of a civil action in the Courts.

This process allows DIA to leverage existing relationships with online content hosts and to use the most appropriate non regulatory/regulatory levers, depending on the situation. This helps to mitigate the risk of over-reaction to regulation by technology industry groups.

In relation to investigation of mechanisms for filtering or blocking objectionable content online, under the proposed amendment DIA would have authority to lead investigation of such mechanisms. DIA will work with relevant agencies, industry groups and experts to develop options for such mechanisms, which would need to be set out in regulations.

Major online content hosts have generally indicated a willingness to be 'good corporate citizens' and comply with the legal obligations of the jurisdictions that they operate within.

As outlined in section 7.1, DIA will be involved in ongoing monitoring of the effectiveness of these changes and may consider further amendments as part of the broader media content regulation system review.

6.2 What are the implementation risks?

The preferred option of a suite of changes has some manageable implementation risks and some unknown uncertainties that can be monitored, and mitigations developed in response. Implementation of the proposals relies on the assumptions that:

- large ISPs and online content hosts are willing and cooperative partners;
- political and public support for the changes; and
- sufficient operational resourcing to implement the proposals.

Investigation and implementation through regulations of mechanisms to filter or block objectionable content online (as would be provided for by the amendment to introduce statutory authority), may be controversial and could raise constitutional and legal issues that will require appropriate consideration of trade-offs. To mitigate risks of ineffectiveness, opposition, and duplication, it will be critical to engage with:

- industry experts at the design stage so that options are effective and able to be operationalised; and
- officials from relevant agencies to ensure rights of freedom of expression are not infringed and other relevant constitutional and legal issues are given due consideration; and that work is complementary rather than duplicative of other censorship and intelligence efforts in New Zealand.

The CVE Ministerial group will provide a mechanism for escalating issues, as well as ensuring consistency with the Christchurch Call and other related government initiatives.

This programme of change must be underpinned by a partnership approach with industry progressing in parallel and continuing as part of the broader media regulation review, to foster productive partnerships at both a national and international level. Any potential future clashes with broad media regulatory regime review will be considered at a future date and will be informed by the specifics/context of the potential clash.

Section 7: Monitoring, evaluation and review

7.1 How will the impact of the new arrangements be monitored?

Extension of standard monitoring processes to cover new functions

The OFLC currently monitors and reports on classification complaints and will implement operational uplift following the approved additional funding to support CVE efforts. OFLC and DIA could also develop an ongoing programme of public engagement and research on the efficacy of the initiatives, with the results reported and published.

Officials will provide updates as required to the newly established CVE Ministerial Group. Updates on key milestones will also be provided as part of the CVE Ministerial Group work plan.

DIA will establish a reporting regime to the Minister of Internal Affairs (and other authorities as identified as appropriate) to ensure Government uses the new filtering/content take-down mechanisms appropriately. DIA will create and publish reports on the DIA website (most likely annually), which would likely provide information on how often the mechanisms are used and what type of content has been blocked. The reporting regime would provide confidence that the mechanisms are not being overused/are being used for their intended purpose. Should a similar event to the Terror Attacks happen again, any evaluation of the response would include examination of the effectiveness of the take-down powers.

A similar reporting function would be used to track OFLC's interim decision ability – tracking the number of interim decisions made and whether they align with the Chief Censor's final decision.

Monitoring and review would include increasing the level of data collection on the quantity of violent extremist and terrorist online content, which is difficult to establish given the nature of the internet and that content continuously changes.

Assessment of policy effectiveness

It will be valuable to articulate from the point of introduction of legislative changes the desired outcome of changes, and a way to test whether the outcomes have been achieved (or are on track for achievement). Specifying desired outcomes and the data required to measure these from implementation outset is preferable.

The scope of this outcomes framework could be limited to DIA-led policy changes, or it could be designed to complement and/or encompass the wider and interconnected Government work programme on CVE.

Work to design an outcomes framework and monitoring approach could be undertaken in parallel to preparing for implementation of policy changes (i.e. from approximately March 2020). Lessons learned in working with ISPs, and the interaction with the Chief Censor's office will be applied and the system will be improved over time.

7.2 When and how will the new arrangements be reviewed?

The short-term effect of these policy proposals will be reviewed as part of the broader media regulation review, terms of reference for which are intended to be considered by Cabinet in 2020. This broader review provides an opportunity to consider any initial results from policy interventions, and if desired, to test whether any other options or approaches should be considered to further enable an improved response to countering violent extremist content online. Stakeholders will have an opportunity to raise their views during the consultation for the broader reform.

Section 8: Glossary

8.1 Key terms used throughout this Regulatory Impact Assessment

- **Harm** is defined as "behaviour online which may hurt a person physically or emotionally. It could be harmful information that is posted online, or information sent to a person."[16]
- **ISPs and online content hosts** refer to:
  - internet service providers, means companies that provide subscribers with access to the internet, such as Spark and Vodafone; and
  - online content hosts as defined in the HDC Act, i.e. "in relation to a digital communication, means the person who has control over the part of the electronic retrieval system, such as a website or an online application, on which the communication is posted and accessible by the user."[17]
- **Objectionable** is as defined in the FVPC Act, i.e. including content which "describes, depicts or expresses, or otherwise deals with matters such as sex, horror, crime, cruelty or violence in such a manner that the availability of the publication is likely to be injurious to the public good". In particular, the analysis relates to content that could be classified as objectionable under sections 3(2)(f) and 3(3)(d) of the FVPC Act, which most directly relate to violent extremism.
- **Violent extremism** is defined by New Zealand's Ministry of Foreign Affairs and Trade as "the justification of violence with the aim of radically changing the nature of government, religion or society".[18] This definition is closely aligned with the definition of a "terrorist act" set out in the Terrorism Suppression Act 2002: "a terrorist act is "an act

[16] U.K. Department for Digital, Culture, Media & Sport (2019): Online Harms White Paper.

[17] Subject to PCO advice, it is intended that where possible and appropriate, legislative amendments use the definition of IPAP or internet protocol address provider, as per section 122A of the Copyright Act 1994 to cover ISPs. The HDC Act uses this terminology rather than ISPs.

[18] [Redacted under S9(2)(f)(iv)]

[…] [that is] is carried out for the purpose of advancing an ideological, political, or religious cause", and with the intention of "[inducing] terror in a civilian population […]"[19]

[19] Section 5(1) of the Terrorism Suppression Act 2002