

Interim Regulatory Impact Statement: A Modernised Regulatory Framework for Media and Online Content in New Zealand

Coversheet

Purpose of Document	
Decision sought:	Release of a public discussion document seeking views on the Department of Internal Affairs' preferred approach to building a framework for regulating media and online content in New Zealand.
Advising agencies:	Department of Internal Affairs
Proposing Ministers:	Minister of Internal Affairs
Date finalised:	17 August 2022
Problem Definition	
<p>New Zealanders are experiencing a wide range of unacceptable harms from media and online content that our existing regulatory mechanisms are unable to address or mitigate</p> <p>Changes in the way we rely on and interact with media and online content are associated with a wide range of potentially harmful experiences from interaction with content, particularly on online platforms. These harms can be experienced by individuals, communities, and society as a whole. The harms being experienced through interaction with media and online content include unacceptable impacts on the wellbeing of New Zealanders. While it is not possible to completely eliminate content-related harm, and content can often be one among many contributing factors to wider societal issues, the fragmented and outdated nature of our existing regulatory mechanisms in New Zealand mean that harms caused by content are not being addressed or mitigated as effectively as they could be and in line with New Zealanders' expectations.</p>	
Executive Summary	
<p>The communication of information, ideas, opinions, stories and pictures through media and online content is a critical element of the relationships between individuals, and between individuals and society. Media and online platforms, and the content available on them, play an integral role in New Zealand and the world we live in today. They deliver educational, informational, recreational and developmental information, raise awareness of global news, educate people on new subjects, deliver entertainment services and connect businesses, friends and family.</p> <p>The way that media services are provided is evolving rapidly, which is having a profound effect in the way we interact with and rely on media and online content. With the development of the Internet, content has become increasingly accessible, immediate, prevalent, and has become a constant part of our lives. At the same time, millions of</p>	

people have been empowered to create and “publish” content on social media, for a mass audience.

This unprecedented growth in the availability and sources of content has also exaggerated existing, and created new, risks of harm to individuals, communities and society. Just as interaction with content has become a regular and indispensable part of everyday life, the dangers of everyday life have become increasingly evident in the content that is available today. Media and online content has become a distorting mirror of society – reflecting, and sometimes twisting and magnifying, negative aspects along with the good.

Governments in New Zealand, like in many other countries, have historically sought to protect the public interest by regulating or prohibiting the public’s access to content seen as threatening that interest. Approaches to content regulation have changed as changes in the nature and role of media, and changing social attitudes, have changed. However, regulation has seldom managed to keep pace. Our current set of regulatory mechanisms is fragmented, incomplete, and developed over 30 years ago, which means that we are ill-equipped to face major changes in the role of media and online content and the positive and negative impacts that this can have on individuals and society.

The Content Regulatory Review (the Review) is seeking to address this problem by developing a broad framework to guide the regulation of media and online content in New Zealand. This will minimise the risks of harm from interacting with content while preserving important freedoms and the benefits that interaction brings.

Targeted engagement was held from September 2021 to July 2022.

- Wide-ranging engagement was held with agency, community and industry stakeholders from September 2021 to April 2022 to understand the harm that individuals and communities are experiencing from media and online content and aspirations for how this can be addressed;
- We then tested the feasibility of possible regulatory and non-regulatory tools with stakeholders that have existing regulatory responsibilities, regulatory design expertise, or other relevant expertise including in rights frameworks. This was done to gauge how well the regulatory system would operate as whole to address harms, while maintaining existing standards for freedom of expression and freedom of the press, and consideration about how Te Tiriti o Waitangi principles and te ao Māori could be embedded within the regulatory framework.
- The feedback received have informed the levers that could be used in the new framework. More details on the levers have been outlined in Section 2 of the RIA.

Following the targeted engagement process, we are initiating public consultation on a discussion document setting out a proposed approach for a new content regulatory framework that lifts the role for industry through a co-regulatory model. This co-regulatory approach recognises that responsibility for minimising harms must be shared between Government, Media Service Providers¹ (MSPs), and individuals and communities.

¹ Media Service Providers: refers to all entities who provide, or enable access to, content to the public. This means any company or entity that provides access to content to the public, irrespective of whether the content is created, developed, edited or curated by or on behalf of the entity, and access to the content is provided free of charge or other consideration.

Regulatory change is needed as the current system operates under existing legislation that will need to be amended, repealed or replaced. Consideration of non-regulatory tools were considered within this overarching co-regulatory approach.

Three options were considered in relation to the role of Government within this co-regulatory approach. The resulting variations on the co-regulatory mode are:

- *Supportive approach* – takes a primary focus on government supporting and encouraging industry and community initiatives to achieve the objectives.
- *Balanced approach* – takes a balanced mix of government regulatory intervention and a focus on partnering with industry and community stakeholders to achieve the objectives.
- *Prescriptive approach* – the primary focus of this approach is on government regulatory intervention to achieve the objectives.

These options were assessed against the status quo in relation to nine objectives identified as necessary for an effective harm minimisation framework. The assessment indicated that for each objective, a balanced co-regulatory approach that would see Government establish legislative requirements to meet minimum expectations, while supporting industry and community implementation and innovation is the option that best meets the criteria. While all three options are described in the discussion document, the balanced approach is identified as the Department's preferred approach.

Detailed development of the preferred approach, including institutional arrangements, legislative content and implementation timeframes, will not be undertaken until after feedback on the discussion document has been received and assessed. Detailed assessment of the marginal costs and benefits of the preferred approach cannot therefore be undertaken at this time. However, it is expected that the balanced approach will significantly increase media safety benefits to individual New Zealanders, communities and society. This will be in exchange for a moderate initial investment and operating commitment from Government, and the lowest possible costs to media and online sectors that are consistent with the Government's public safety objectives.

Limitations and Constraints on Analysis

Scope and range of options considered

Cabinet agreed to initiate a broad, harm minimisation-focused review in May 2021. This harm minimisation focus of the Review has constrained its scope, as it covers matters dealt with by existing legislative frameworks that represent this objective in relation to media content², and matters that equivalent modern and fit-for-purpose frameworks might address. The focus on the risk of harm from media and online content itself has also meant that regulatory frameworks that deal primarily with harmful behaviour that happens to occur online (e.g., online fraud and scams, harmful private communication between individuals) are also peripheral to the Review. Relationships and interdependencies with those frameworks will, however, require ongoing attention as the Review progresses.

In addition, it is important to recognise that attitudes such as racism and sexism, extremist ideologies, and distrust of government are symptoms of broader and complex social

² Principally the Films, Videos, and Publications Classification Act 1993 (the Classification Act) and aspects of the Broadcasting Act 1989.

problems, of which content is only one contributing factor. As such, the scope and range of options considered are focused on how content itself can have an impact on or lead to these harmful behaviours.

Timeframes

The nature and extent of research and public engagement possible in the Review has been determined by timeframe expectations set initially by Cabinet and modified by subsequent Ministerial decisions. The Review is currently working to a deadline of final policy decision by mid-2023, which limits possible public consultation on proposals to the discussion document on high level approaches that is the subject of this analysis. Targeted engagement on detailed proposals will be undertaken with relevant stakeholders as these are developed following public consultation.

Early targeted engagement with industry, experts, and community and population groups was undertaken to inform the Review's understanding of the nature and extent of harm experienced in New Zealand as part of the development of the problem statement. The COVID-19 Delta outbreak in August 2021 caused delays to the targeted engagement process, and engagements became limited to a virtual-only format. This, in addition to consultation fatigue felt by some communities, and time and resource-specific demands (e.g. COVID-19 community vaccination efforts) during the targeted engagement period may have led to lower levels of participation in some engagement sessions. In turn, this may have impacted on the range and diversity of opinions that could have been canvassed, and the robustness of group discussions, that could otherwise have been achieved.

Assumptions

The development of the objectives and assessment of the options necessarily reflect a number of fundamental assumptions. These include:

- that people will be motivated to avoid or manage their interaction with content that they know may be harmful to them;
- that not all negative reactions to content justify a regulatory response, and that subjective experiences of discomfort or offence are unavoidable risks in a free and diverse society that consumers should be enabled to manage themselves;
- that the right to freedom of expression does not extend to a right to detract from the rights and freedoms of others, or to freedom from consequences;
- that MSPs generally have commercial incentives to maximise audience engagement and, in the case of social media platforms, to enable immediate and unimpeded posting of user-generated content;
- that there is a role for Government stewardship and regulatory oversight in setting standards for reducing harm;
- that alignment of approaches with comparable democratic jurisdictions will support attainment of the objectives of the Review by:
 - enabling coordination, cooperation and shared experience in pursuing shared objectives;
 - enabling joint enforcement and information initiatives across national borders; and
 - maximising consistency of expectations imposed on global platforms.

Limitations on data

Timeframe limitations, together with the complex and evolving context for the Review, have also impacted on the nature and extent of data accessed and research commissioned. While research to understand and compare approaches to the regulation of media and online content in other jurisdictions has been undertaken, research to better quantify the impact of media harm in New Zealand is very patchy. A systematic overview of existing data relating to harm experienced by New Zealanders from content is being developed and will be available to support public consultation and subsequent development of detailed proposals. In the meantime, research undertaken by, or for organisations such as Internet NZ, Netsafe and Te Mana Whakaatu Classification Office has proved valuable.

Levels of confidence

Despite these constraints, Ministers can have confidence in the analysis and selection of the balanced approach as the Department's preferred option for public engagement. That approach provides considerable flexibility to ensure that appropriate balances can be struck between effective promotion of the harm minimisation objectives, minimising compliance costs to media and online platforms, and any limitation or erosion of access to media or online content for the public.

Approved by

Suzanne Doig
General Manager, Policy Group
Department of Internal Affairs



22 August 2022

Quality Assurance (completed by QA panel)

Reviewing Agency:	Department of Internal Affairs
Panel Assessment & Comment:	<p>The Department's Regulatory Impact Analysis panel (the panel) has reviewed the <i>Interim Regulatory Impact Statement: Proposed Framework for Regulating Media and Online Content in New Zealand</i> (the interim RIS) in accordance with the quality assurance criteria set out in the CabGuide.</p> <p>The panel also reviewed the inter-agency consultation version of the draft Discussion Document <i>A proposed new framework for regulating media and online content in Aotearoa</i> (draft as at 20 July 2022).</p> <p>The panel members for this review were:</p> <ul style="list-style-type: none">• Amanda Shaw, Principal Policy Analyst (Chair)• Alan Edwards, Senior Policy Analyst (Policy member)• Leeza Boyd, Senior Policy Analyst (Policy member)

- Grace Turner, Policy Analyst (Policy member)
- Fergus Campbell and Dana Visnovsky, Policy Analysts (Secretariat)

The panel considers that:

- the draft Discussion Document is likely to lead to effective consultation and support delivery of RIA to inform subsequent decisions; and
- the information and analysis summarised in the interim RIS meets the quality assurance criteria.

The draft Discussion Document outlines three alternative options to the status quo, with a stated preference for one of the options. The draft Discussion Document identifies that the three options are on spectrum with some fluidity between the options. The draft Discussion Document provides an opportunity for feedback outside the preferred approach.

The interim RIS and draft Discussion Document together provide a detailed overview of the status quo and demonstrate a clear understanding of the problem to be solved. The interim RIS outlines that a good level of early engagement with a range of stakeholders supports the conclusion on a preferred option for the purposes of the Discussion Document.

The high level nature of the policy options considered means that at times the distinctions between the options, and between each option and the status quo, is a little blurred. Therefore, the regulatory impact of each of the options cannot be clearly defined. This should be able to be addressed in subsequent RISs as the measures, regulatory interventions, and sanctions become more concrete through further policy development.

The panel considered that the format of the analysis section in the interim RIS could have been more concise. The multiple-objective approach makes it more difficult for the reader to understand the relative advantages and disadvantages of the 4 options. However, we do not think that this detracts from the quality of the analysis, and it does provide extra detail that may be useful for some submitters.

Section 1: Diagnosing the policy problem

What is the context behind the policy problem and how is the status quo expected to develop?

The Status Quo

1. The communication of information, ideas, opinions, stories and pictures through media and online content is a critical element of the relationships between individuals, and between individuals and society. Media and online content (content)³, and those that make content available (MSPs) play a critical role in spreading news, keeping people updated on important events, and raising awareness, and influencing our opinions about the world around us. Content, and those who provide it, have played an integral role in shaping the world we live in today, from maintaining democracy to promoting or curtailing social movements.
2. This is as true in New Zealand as in any other society. Historically, a wide range of traditional media outlets have developed and succeeded at both national and provincial level. However, like much of the world, New Zealand is experiencing a number of major changes in the way that media and online content services are provided, which together are having a profound effect in the way we interact with and rely on this content.

Digitalisation has changed the way content can be accessed

3. Technological changes have greatly extended the ability of content to reach a mass audience, starting with the development of industrial printing presses, through the evolution of radio and television broadcasting, to dissemination of digital content over the internet. Not only has this increased the access of content to huge audiences, but it has also greatly increased the number and range of media outlets that consumers can access. It has also greatly increased the degree of interactive control that audiences have over what content they experience and when they experience it.

The creation and dissemination of content is experiencing a “democratisation”

4. The same technological changes have shifted the ability to create and “publish” content from relatively few media companies to millions of users on social media platforms. Almost anybody can now make their music, art, observations, or interests accessible to millions of others without needing the prior approval or curatorial oversight of a publisher or editor.
5. Before the internet, content was disseminated via a limited number of media channels – literature and news publishers, television and radio broadcasters, and cinemas. Consequently, content delivery was controlled by a limited number of agencies due to economic and technological factors – the number of television or radio stations available was limited and significant economic resources were required to publish a book or print a newspaper. However, the internet has created an environment in which disseminating content requires far fewer inputs or resources and has opened content dissemination, and creation, to the public.

Media services have gradually become more commercialised

6. While newspapers have historically always been privately owned, broadcasting media were commonly developed and provided as a public service by Governments. Pressures on Government revenues, an increasing adoption of commercial management and business models, and increasing market competition from the private

³ *Media and online content (content)* refers to: any communicated material (for example: video, audio, images and text) that is publicly available, regardless of how it is communicated.

sector have led to a gradual commercialisation of media services. Most media services are now predominantly reliant on advertising revenue to fund their services. The diversification of media services that has accompanied the proliferation of digital and internet services has increased the dominance of private sector interests in content provision.

Globalisation means content, and those that provide it, now transcend national boundaries

7. The rise of online media has also enabled the rise of large global media companies whose markets and products reach across borders. Because the success of many online applications – particularly social media – is dependent on number of users, market share in this space has become dominated by a small number of very big entities.⁴

The internet and continued technological developments have led to a digital convergence

8. The development of internet technology has also led to convergence, a term used to describe the common delivery of previously discrete service functions such as broadcasting and telecommunications over shared digital platforms. Convergence has reduced the boundaries between previously distinct industries. For example, news publishers that used to produce and distribute newspapers to deliver news content now use the internet for the same purpose. Convergence offers several benefits to consumers and businesses, including improved access to markets for content creators and reduced entry barriers for local and international industry players that seek to enter the New Zealand market. This is disrupting traditional business models by enabling greater competition and innovation in industry revenue models and product offerings.
9. Convergence has also allowed international content providers readier access to New Zealand audiences, and vice versa. Regulators in New Zealand have always had to consider foreign-sourced content, but this has expanded far beyond literature, films, and, later, video games. New Zealanders can now access content from around the world, with the main preconditions being an internet connection and access to a device. This increased ease of access has exposed New Zealand audiences to greater quantity and wider diversity of potentially harmful content.
10. Convergence and internationally sourced content have also exacerbated the risk of harm from mis and disinformation.⁵ The association of this content with content from trusted sources gives it greater credence, as can the international context. This risk of harm caused by the spread of mis/disinformation on the internet (e.g., misinformation relating to COVID-19) is exacerbated by the lack of regulatory responses in New Zealand to address this issue.

New Zealanders are online now more than ever

11. According to online research about New Zealanders' attitudes to the internet⁶, 93% of participants use the internet at least once a day at home, and 66% use it at work. Social media is frequently used with at least 79% of participants in the online survey using at least one of the four Meta platforms⁷ daily. The number of home fibre connections to the internet also continues to increase with 62% of participants' homes

⁴ Commonly known as Big Tech or the Tech Giants, these entities are Alphabet (Google and YouTube), Amazon, Apple, Meta (formerly Facebook), and Microsoft.

⁵ Misinformation is false information that is spread regardless of whether there is intent to mislead. Disinformation is deliberately misleading or biased information, including manipulated narrative and facts.

⁶ <https://internetnz.nz/new-zealands-internet-insights/new-zealands-internet-insights-2021/>

⁷ This refers to Facebook, Instagram, Messenger and WhatsApp

having a fibre connection in 2021. This means more people will have access to a fast and stable connection to the internet, making accessing online content easier.

There are many benefits to increased accessibility of media and online content

12. Interaction with media and online content is an increasingly important part of our lives. It delivers educational, informational, recreational and developmental information; from raising awareness of global news, educating people on new subjects, delivering entertainment services such as films and television shows, and connecting businesses, friends and family.
13. While news media and journalism have always had an important role as the fourth estate – serving as a check on government – convergence has made it easier for a much wider range of people (e.g., bloggers, citizen journalists) to perform journalistic functions. It has increased opportunities for political and social debate and the contest of ideas. In addition, the increased accessibility of content brought on by digital convergence and the internet, has enhanced the freedom of expression⁸ of many, which in turn supports a healthy and vibrant democracy.
14. Overall, content has become more accessible, diverse, and prevalent with the development of the internet. Unlike previous decades, real-time information can be sought by nearly anyone, anywhere, at any time. This proliferation of information has enabled profound opportunities for business, social groups, families, and individuals. It has improved economic, health and social outcomes for many, enabled opportunities for individual and company growth, connected people and various groups, and contributed to a more globalised society.
15. This is reinforced in a July 2021 survey by Netsafe, which found that a majority of people who participated in the research continue to think that the positives of the internet outweigh the negatives – 71% of respondents thought the internet and digital technologies have positively impacted them.⁹
16. In addition, a more accessible content environment has also contributed to supporting the revitalisation of te reo Māori, with a 2017 survey finding that 59% of Māori respondents (52% being Māori youth) indicating that they use the internet or social media to keep in touch with Māori culture.¹⁰

However, there are also many existing and emerging harms from content that New Zealanders are increasingly concerned about

17. This unprecedented growth in the availability of media and online content has also exaggerated existing, and birthed new, harms (examples of these harms are outlined in detail in the next section). Not only has interaction with content become a regular and inseparable part of everyday life and behaviour, but everyday behaviour has become evident in the content that is experienced. Media and online content have become a distorting mirror of society, reflecting and sometimes magnifying the bad and divisive features along with the good.
18. Not only does the explosion and accessibility of content pose risks to population and societal wellbeing, but harmful content has also become increasingly prevalent. The ongoing evolution of digital media and the rise of new content platforms has resulted in

⁸ Freedom of expression, which includes the freedom to seek, receive, and impart information and opinions of any kind in any form, is affirmed in domestic law through the New Zealand Bill of Rights Act 1990 and through international conventions and commitments.

⁹ The survey comprises data from 808 people in New Zealand aged over 18. It sampled people from all regions, ethnicities, religions and age groups. The full survey is available here: <https://www.netsafe.org.nz/wp-content/uploads/2021/07/Netsafe-State-of-the-Online-Nation-Study-2021.pdf>

¹⁰ <https://www.tpk.govt.nz/docs/tpk-maihi-karauna-measuring-the-current-state.pdf>

a significant increase in the potential for New Zealanders to be exposed to harmful content or be harmed by consistent exposure to content.

19. Previously, individuals either proactively chose what content they experienced (for example DVDs, print media) or consumed broadcast media by selecting from a limited number of predetermined schedules (broadcast radio and television). In contrast, consumers of online media, including social media, can choose from an infinite range of sources of information and entertainment that can be accessed at any time.
20. In order to facilitate and guide consumer navigation of the limitless amount of content available online today, social media companies and search engines use algorithmic recommendation systems to narrow the range of content visible to consumers. These recommendation systems are usually individualised (i.e., based on data generated by the individual and their contacts) and are almost always invisible to the consumer. They aim to maximise consumer engagement by leading them to the type of content the algorithm predicts that they will most likely engage with or react to. This means some individuals are shown a narrow range of content without knowing they are experiencing individualised ‘filter bubbles’ or ‘echo chambers.’ This creates the illusion that the type of content they are engaging with is more prevalent and dominant than it is. This can also create or contribute to polarising attitudes among different groups of people, which can erode social cohesion and community wellbeing. At its worst, it can also contribute to marginalisation, radicalisation and extremism, including terrorism or violent extremism.
21. People have different degrees of vulnerability to harmful content and different levels of confidence in navigating media and online spaces. This means that some people will not be directly harmed by consuming the same content that might harm others. In addition, when people engage with content that they are vulnerable to, this can drive behaviour that harms themselves and others in both the online and offline worlds.
22. For example, data from the 2021 General Societal Survey, which provides new measures of wellbeing such as sense of control and feeling safe using the internet, found that 41% of disabled people reported low ratings of sense of control compared to 24% of non-disabled people. Disabled people are also less likely to feel safe using the internet for online transactions than non-disabled people.¹¹
23. New Zealand society has become increasingly diverse, and we see this through an increase in cultural and religious diversity, as well diversity in people’s identities as individuals.¹² Currently, the way content is regulated does not necessarily reflect the changing societal expectations of what type of content could be considered harmful.

Evidence of harms in New Zealand – presenting symptoms and adverse outcomes

24. With the ease of access to online content, New Zealanders are also becoming increasingly concerned about content found online. The July 2021 Netsafe study found that 68% of respondents believe the internet is more dangerous than it was five years ago.¹³ Concerns are related to:
 - misleading and damaging content such as misinformation, disinformation, online conspiracy theories, extremist material, and racist or discriminatory material;

¹¹ <https://www.stats.govt.nz/information-releases/wellbeing-statistics-2021/>

¹² [Ethnic group summaries reveal New Zealand's multicultural make-up | Stats NZ](#)

¹³ <https://www.netsafe.org.nz/wp-content/uploads/2021/07/Netsafe-State-of-the-Online-Nation-Study-2021.pdf>

- content that perpetuates racist stereotypes and prejudices against different communities and population groups, which can negatively impact how others view these groups, or damage how these groups view themselves and their identities;
- privacy and security concerns relating to online crime, personal data, identity theft and location tracking;
- people having the agency to control their own data and being able to seek help that is accessible and transparent when things go wrong;
- children accessing inappropriate content; and
- cyberbullying and sharing of dangerous or discriminatory messages.

Individuals, particularly young people, are experiencing experience harm through content

25. Some studies show that the way people consume content, particularly via social media platforms, can have detrimental impacts on their mental health. A 2019 study found that adolescents who spent more than three hours per day on social media may be at a heightened risk for mental health problems, particularly internalising problems such as depressive and anxiety symptoms.¹⁴ There are also suggestions that there is an association between social media use and externalised problems, such as bullying, harassment and attention problems.
26. Similar viewpoints were highlighted in a June 2022 survey of New Zealanders' views conducted by Te Mana Whakaatu Classification Office (the Classification Office).¹⁵ The report showed widespread concern by New Zealanders about harmful content available online, particularly content being seen by children and young people. Some of the findings include concerns about children and young people finding content that depicts sexual violence or harassment, self-harm or suicide, racist or discriminatory comments and behaviours, or violent extremist and terrorist content. More detail on the findings of this report and survey methodology is attached as **Appendix A**.
27. These issues were also highlighted by participants throughout targeted engagement. Participants from across different ages, abilities, ethnicities, gender identities and religions considered children and young people to be most at risk of harm from content, with much of the concern relating to the harm that exposure to inappropriate content can have on both the physical and mental wellbeing of young consumers. Examples raised by participants included the ways in which social media influencers and content can impact young consumers' views on body image, which can lead to unhealthy eating habits, low self-esteem, and poor mental wellbeing.
28. Significantly, young people in our targeted engagement sessions described interaction with content as an intrinsic part of their day to day lives, often inseparable from their 'real' environments. This sentiment is reinforced in a March 2021 survey conducted by the Classification Office, which was a nationally representative survey of more than 2,000 people aged 16 and over, found that 79% of respondents get news or information from social media.¹⁶

¹⁴ <https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2749480>

¹⁵ <https://www.classificationoffice.govt.nz/resources/research/what-were-watching/>

¹⁶ <https://www.classificationoffice.govt.nz/resources/research/the-edge-of-the-infodemic/>

29. It was evident from our youth engagement that young people have a high degree of awareness about social media platforms and the possible harms that they can experience in online spaces. It was emphasised that online issues for young people are largely existing social and personal issues that are carried over into online spaces – issues like low self-confidence and bullying are amplified online. Engagement also found that young people have a 'deep distrust' of most content platforms because they do not trust major social media platforms to protect them from harmful content.
30. In addition to the research and evidence from targeted engagement outlined above, a systematic overview of existing data relating to harm experienced by New Zealanders from content is being developed and will be available to support public consultation and subsequent development of detailed proposals.

Different communities and population groups are experiencing harm through content

31. New Zealand has a diverse and growing range of communities and minority groups such as Māori, rainbow, Pacifica, ethnic, and faith-based communities. These diverse communities have often experienced a range of harms from content, which is amplified by how quickly information is circulated on the internet and the anonymity that the internet enables. For example, community and minority groups can often be unfairly represented or underrepresented in the media, and are targets for hate speech, discrimination, and harassment. In turn, this can contribute to social polarisation, erosion of tolerance for social and cultural differences, and institutional and systemic discrimination.
32. Across all engagements with different community and population groups, it was evident that participants' main concern is social media content and the harms that individuals belonging to these groups – and the groups themselves – can face on social media platforms. Harms identified included hate speech, bullying, harassment, discrimination (racism, gender, ethnicity and faith-based), and misinformation resulting in division between communities. A Netsafe survey found that 16% of Asian participants and 13% of Māori and Pacific participants experienced online hate speech one or more times in the prior year. For New Zealand European or Pākehā respondents, only 9% had experienced hate speech¹⁷. Participants referred to discriminatory content that has impacted them on a personal level and at times on a group level. For example, the issue of misogynistic abuse and violent threats against prominent women in New Zealand – particularly wāhine Māori – was raised a number of times by participants as something that causes harm not only to the women who are the direct targets, but to all women.
33. Community engagement also highlighted that mainstream and news media reporting often conveys an unconscious bias against particular communities and population groups. In addition, distorted views portrayed on mainstream and news media are considered to have a ripple effect on the scale and severity of discriminatory messaging published on social media platforms. For example, many participants who identify as Māori, Pacifica or Asian raised concerns about how mainstream and news media portrayed their respective communities during the initial and subsequent COVID-19 outbreaks and when reporting on vaccine-related information. Participants felt that this caused or exacerbated prejudices towards these communities both online and offline.

Harm experienced online for Māori is exacerbated by historical experience of discrimination and injustice

34. Engagement with Māori highlighted that many of the harms experienced by other communities and population groups are also being experienced by Māori. For Māori,

¹⁷ <https://www.netsafe.org.nz/wp-content/uploads/2019/11/onlinehatespeechsurvey-2018.pdf>.

however, there is an added dimension to harmful content interactions and attempts to address harm, as Māori have historic reasons for mistrusting Government interventions.

35. We also heard that news media reporting across all channels often reflects a bias (conscious or unconscious) against Māori, which is mana-reducing and creates or reinforces negative attitudes toward Māori within society. This has a particularly harmful impact on tamariki and rangatahi, and their relationship with their Māori identity.
36. The Crown-Māori relationship under Te Tiriti o Waitangi places particular attention to understanding and addressing the harms that are being experienced by Māori.

There is an increasing spread of misinformation and disinformation

37. Misinformation and disinformation are not new phenomena; however, the internet has enabled the amplification of the dissemination and impact of this type of content.
38. Misinformation and disinformation have the potential to cause New Zealanders to accept and act on false information, disengage from democratic processes, as well as Government and community programmes, and participate in civil disobedience or active non-compliance. Examples such as the recent protests outside New Zealand Parliament have demonstrated how mis/disinformation that is spread on social media can be a strong contributing factor in influencing some individuals and groups to act disruptively, illegally, and violently to their own and wider society's detriment.
39. Misinformation and disinformation results in a variety of adverse outcomes for society, which can include:
 - growing social polarisation and public frustration;
 - public mistrust of government and traditional media sources; and
 - the rise of alternative information sources and influencer/non-traditional reporting.
40. Eighty-two percent of respondents in a recent Classification Office report reported they felt concerned about mis and disinformation, and 57% of respondents reported seeing false or misleading news or information in the last six months¹⁸. These concerns were also common across targeted engagement. Many participants shared their concerns and experiences about harms that they, their whānau, or people they know have experienced because of misinformed views expressed on both mainstream and online media, which is then repeated through social media platforms such as Facebook.

The current regulatory situation

41. The regulatory mechanisms and tools that currently exist in New Zealand are fragmented as they were developed independently of each other and around the type of media or format by which the content is made available (e.g., through linear television, cinematic release, or radio broadcast). Fragmented regulatory mechanisms and tools also mean that responsibilities over the same content can often fall under the purview of a number of agencies and regulatory bodies.

¹⁸ https://www.classificationoffice.govt.nz/media/documents/The_Edge_of_the_Infodemic.pdf

42. The following agencies are primarily responsible for regulating content in New Zealand. These organisations are:

Authorised Agency	Domain	Source of powers
Office of Film and Literature Classification (OFLC)	Films, videos, images and literature on any media; also defines what content is illegal or 'objectionable' ¹⁹	Films, Videos, and Publications Classification Act 1993
Department of Internal Affairs	Enforcement action on illegal content, primarily online	State Sector Act 1988
Broadcasting Standards Authority (BSA)	Radio and television broadcast content, including some online content from these providers	Broadcasting Act 1989
Ministry of Justice (MoJ) with Netsafe as current approved regulator	Provides remedies serious or repeated harmful digital communications	Harmful Digital Communications Act 2015
Advertising Standards Authority (ASA)	All advertising, on any media	Self-regulatory body (voluntary basis)
Media Council	Editorial and video-on-demand content published by members	Self-regulatory body (voluntary basis)

43. New Zealand Police also deals with criminal content, and there are other bodies (such as the Computer Emergency Response Team or CERT, Netsafe, and Internet NZ) that operate on the periphery of the New Zealand 'content regulatory system'. The main organisations making up the regime and the functions they exercise are:

Function	Organisation
Enforcement	DIA, Police, Customs
Classification and Labelling	OFLC, Film and Video Labelling Body
Standards Enforcement and Complaints Investigation	BSA, ASA, Media Council, Domain Name Commission NZ

44. The fragmented nature of existing regulatory mechanisms and the shared, overlapping regulatory responsibilities of different agencies mean that current approaches to content regulation cannot adapt to the ongoing evolution of content and the platforms on which this occurs. This is particularly the case, as the same content is now often disseminated across different media types at the same time. Inconsistent standards are often applied to the same content depending upon the type of media or format. It is not clear what type, scale and trends of harm occurring from legal but harmful content because different agencies can hold different parts of related information.

¹⁹ Under the Classification 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.

45. In addition, existing mechanisms and tools do not adequately cover online content as they were developed before the widespread use of the internet. Newer forms of media, such as social media, are not subject to any government regulation in New Zealand. This means that legal content on online platforms is largely self-regulated.
46. While the BSA and the Media Council do have some authority over certain online content, most online content is outside of their regulatory scopes. In addition, MoJ is responsible for the HDCA with Netsafe being the current approved agency to investigate complaints. The HDCA provides remedies for certain online behaviours, regardless of whether those take place through publicly available content²⁰ or in private²¹. Furthermore, while the Films, Videos, and Publications Classification Act 1993 (the Classification Act) applies to a certain extent, it does not provide full coverage for all content. The resulting regulatory regime is incoherent in the digital age and is characterised by its large gaps with respect to online content.

Other related initiatives in New Zealand

47. Partially because New Zealand lacks a cohesive content regulatory system, there are several other related regulatory or quasi-regulatory initiatives that aim to address, at least to some degree, harm caused by content. In particular, the aftermath of the 15 March 2019 terrorist attacks on Christchurch masjidain has led to a number of Government initiatives to address both harmful online content, as well as related offline behaviours and interactions that may cause harm.
48. Following the 15 March 2019 terrorist attacks, the New Zealand Government, together with France, brought together Heads of State, Governments and leaders from the tech sector to adopt the Christchurch Call. The Call is a commitment by Government and tech companies to eliminate terrorist and violent extremist content online, while upholding the right to freedom of expression.
49. The Government also recently passed laws that directly addressed a gap in the classification regime with regard to livestreaming, to give the Government and Chief Censor powers to act more swiftly when violent extremist material is published online. These new powers were recently used in response to the Buffalo New York shooting in May 2022, where the offender's manifesto was given an interim assessment by the Chief Censor of being an objectionable publication within two hours of the incident.
50. The New Zealand Government also announced a Royal Commission of Inquiry (RCOI) into the attacks, which investigated what measures agencies should take to prevent future terrorist attacks, like that in Christchurch. The RCOI report was released in December 2020 and made 44 recommendations covering both national security and wider social and community matters.
51. In response, the Government has a number of initiatives underway that either directly implement the Inquiry's recommendations or support a broader Government response to addressing online harm. These include:
 - A significant programme of work to strengthen social cohesion. The social cohesion work programme included the development of a strategic framework, a measurement framework, a government work programme, and information sheets for communities and sectors to contribute to and strengthen social cohesion in New Zealand. The work programme is now moving into implementation.

²⁰ This can include threatening posts on a public platform or non-consensual image-based sexual abuse (sometimes referred to as 'revenge porn').

²¹ This can include harassment by private messaging or email.

- Proposed law changes to strengthen protections against speech that incites hatred and discrimination. These proposals aim to increase the groups of people that are protected by the incitement provisions and clarify the protections for groups against wider discrimination.
 - Proposed changes to the education curriculum and wider education initiatives aimed at improving New Zealanders' understanding of ethnic, cultural and religious diversity as well as build emotional and social resilience; along with the following initiatives:
 - i. provided guidance to schools to develop safe digital technology strategies and practices in the best interests of students;
 - ii. secured additional funding to refresh and enhance Positive Behaviour for Learning (PB4L) delivery through Early Childhood Education and Schooling, including Māori Medium Education; and
 - iii. recently introduced regulations to ensure that tertiary sector providers have systems and resources to reduce bullying, discrimination and harm and support learners' inclusion, wellbeing and safety.
 - The establishment of the Ministry for Ethnic Communities. The new Ministry replaced the Office of Ethnic Communities to provide an increase in visibility and mana for ethnic communities, to lift cultural competency across government institutions and ensures that services can be tailored to meet the needs of ethnic communities, to provide greater ability to deliver on Government's priorities, and to establish and maintain strong relationships with New Zealand's ethnic communities.
 - The establishment of the New Zealand Police's Te Raranga programme. The programme is a victim-centric approach to hate crime, aiming to develop resources that make it easier for victims and their families to report hate crime, public education to prevent hate crimes, and improvements to the Police's systems, processes and frontline practices to identify, record, manage and respond to hate crime.
52. In addition to Government initiatives, there have been a range of voluntary/non-regulatory initiatives over the years that have attempted to address the immense changes brought on by the rise of the internet and online platforms, in the absence of Government regulation. For example, both the New Zealand Media Council and Advertising Standards Authority, which operate voluntary self-regulatory schemes, have extended their remits to deal with content in a broader range of circumstances.
53. More recently, Netsafe has led the development and launch of the Aotearoa New Zealand Code of Practice for Online Safety and Harms. This voluntary Code has an overarching aim of building and maintaining a safer online environment by setting out a series of objectives that signatories must report on, and commit to, to maintain their signatory status. The Code proposes that signatories, predominantly digital platforms, commit to a set of guiding principles, outcomes, and measures that promote online safety, which relate to child sexual exploitation and abuse, bullying or harassment, hate speech, incitement of violence, violent or graphic content, misinformation and disinformation.

International Initiatives

54. Like New Zealand, other countries are grappling with outdated content regulatory systems that do not have the flexibility to respond to new forms of media and the risk of harm they pose. Examining international work in this area and combining it with our own New Zealand experience, including community engagement, has highlighted that harm is highly complex and dynamic.
55. The United Kingdom, Canada, Ireland, and Australia, as well as many other jurisdictions, are in the process of updating their content regulatory systems. An overview of international approaches to the regulation of content is attached at **Appendix B**. Any regulatory response in New Zealand must be tailored to our society and its needs. Nevertheless, aligning with other jurisdictions, where appropriate, provides the opportunity to maximise the efficacy of any new regulatory system in three ways:
 - enabling coordination, cooperation and sharing of experience in pursuing shared objectives;
 - enabling joint enforcement and information initiatives across national borders; and
 - maximising consistency in the expectations placed on global platforms.
56. In addition, internationally, technology and social media platforms have progressively shifted to an industry codes-based approach to moderating content for harm. Examples include the EU Code of Practice on Disinformation, the EU Code of Conduct on Countering Illegal Hate Speech Online, the Australian Code of Practice on Disinformation and Misinformation and the Digital Trust & Safety Partnership Best Practice Framework. In July 2022, Meta, Google, TikTok, Amazon and Twitter signed up to the Aotearoa New Zealand Code of Practice for Online Safety and Harms alongside Netsafe and NZ Tech, to actively reduce harmful content on their platforms.

How the status quo is expected to develop if no action

57. If the current situation continues, the fragmented regulatory system will remain and gaps (i.e., content that is not comprehensively addressed by any regulation) are expected to grow. Feedback from community engagement has confirmed that consumers are confused by New Zealand's current regulatory arrangements in terms of what requirements apply in what circumstances and, particularly, what avenues for redress may exist in any specific instance. This confusion will only be exacerbated over time if there is no regulatory change that comprehensively aligns and simplifies the content regulatory framework.
58. Continuing to address identified regulatory gaps by ad hoc additions to the current outdated regimes will only exacerbate the already fragmented nature of New Zealand's content regulatory coverage. A non-cohesive regulatory response will mean consumers remain confused about how they can address content harms.
59. While the media and online industry have made incremental changes over the years to their practices in an effort to address regulatory gaps, these have been reactive, ad hoc and discrete. Continuing to respond in this way is likely to lead to greater fragmentation and inconsistency in responses to harms.
60. New Zealand's existing regulatory frameworks capture a shrinking proportion of content and are failing to adapt to changes such as the way in which individuals rely on and interact with content generally.
61. Current regulatory mechanisms also fail to adequately enable people to interact with content safely within existing and accepted thresholds for freedom of expression, and they do not ensure appropriate redress is available when harm does occur. As society, including social values and expectations, will continue to change, it is important for a

modern regulatory system to be adaptable and flexible so both current and future harms can be mitigated.

What is the policy problem or opportunity?

New Zealanders are experiencing a wide range of unacceptable harms from interaction with content that our existing regulatory mechanisms are unable to address or mitigate

62. Changes in the way we rely on and interact with media and online content are associated with a wide range of potentially harmful experiences from interaction with content, particularly through online platforms. These harms can be experienced by individuals, communities, and society as a whole. The harms being experienced through interaction with content include unacceptable impacts on the wellbeing of New Zealanders. New Zealand's existing regulatory mechanisms are unable to effectively address or mitigate these harms, largely because of their fragmented and outdated nature, and their incomplete coverage of content types and channels.
63. The changing role and nature of media in our lives has seen it move from a discrete form of information and entertainment to an integral part of our everyday experience and, particularly, the principal mechanism through which many individuals interact with other individuals and society at large. Just as disputes, social divisions, hate, fear and violence are phenomena that exist within our society, so they are reflected in the content we interact with. According to Netsafe, one in four harmful digital communications were sent as part of an issue that is also taking place offline, while nearly half related to an online issue only.²²
64. Addressing and reducing these phenomena is a critically important social goal, but it cannot be the focus of the Review. What the Review is focussed on are the roles that MSPs and their content play in enabling, amplifying and transmitting harmful impacts to and through the persons that interact with them.

Individuals are experiencing harm through a wide range of content in differing ways

65. A wide range of content has become easier and faster to access due to the digital convergence that is amplified by the internet. As a result of this, individuals are experiencing a wide range of harms in varying ways from the types of content that are being created and the types of content that are being consumed. These harms exist along a spectrum, with the most extreme being child exploitation/abuse content, sexualisation of children, and violent extremist/terrorist content. Harmful content to individuals also includes the non-consensual public posting of sexually explicit images, online harassment of any kind, cyber-bullying and acts of discrimination. At the other end of the harm spectrum, it is important to recognise that subjective reactions of offence or discomfort are likely to be matters for individual consumers to avoid or manage, with the roles for Government and platforms limited to supporting their ability to do this.

The harms being experienced by New Zealanders can be linked to a number of root causes that are not adequately regulated for

66. As discussed above, a wide range of harms is being experienced by New Zealanders as individuals, communities/groups, and at society level. This includes mis/disinformation about a range of public issues, hate speech and discrimination, and negative impacts on the mental health of some individuals associated with their

²² <https://www.netsafe.org.nz/perpetration/>

experiences of social media. The potential for these harms is linked to a range of root causes. As noted above, this analysis is concerned only with those that relate directly to content and how it is experienced.

Imperfect and incomplete information is a root cause of the risk of harm

67. In many cases, the risk of harm from content arises accidentally. While there is ongoing research, the adverse impacts that can arise from interaction with content are not well understood, even in the scientific community. This is also true amongst media providers themselves including those who create and shape media professionally. Consumers of content, and users who create and post content, therefore have even more limited and imperfect knowledge of how harm can arise from interaction with content generally, and with particular types and categories specifically.
68. For consumers and potential consumers, the impact of this knowledge gap is compounded by changes in the way in which their personal experience of content is determined. Historically, consumers proactively selected which print media content they experienced and had almost as much control over which broadcast and recorded content they elected to experience. These choices have been largely supported and informed by a regulatory system that requires potentially harmful content to be identified and labelled as such and regulates (including prohibiting) access to content identified as carrying a high risk of harm.
69. These safeguards apply to only a small proportion of the content experienced today. Consumers have limited ability to anticipate and control the nature of content they interact with, and only a small proportion of potentially harmful content is accompanied by warnings or harm mitigation.

Those who create and publish content have weak incentives to minimise harm

70. Regulatory analysis typically assumes that individuals will act to minimise harm to themselves (or to children and others in their care) if they are aware of the risks. Consumers will, therefore, generally be incentivised to manage the risk of harm to themselves including balancing this against potential benefits from their interaction with media.
71. Those who create and provide content have much weaker incentives to avoid or minimise the risk of harm to consumers, or to others and society through the impact on consumers. What incentives they do face are dependent on their own awareness of potential impacts, coupled with the nature and strength of their concern to avoid harm and the sensitivity (if any) of the markets they operate in to perceptions of safety from harm. Against this, commercial and social incentives to maximise audience numbers and engagement with content may drive the production and publication of content that is potentially harmful to consumers, or can result in harm to others, to communities and to society.
72. Users on some social media can create content and share it with a mass audience without editorial scrutiny (which has historically applied to mass/public communication print and broadcast media) or that the limitations typically associated with communicating in person to a small number of people. The anonymity and reach of social media, coupled with its increasing dominance of individuals' social experience, is significantly weakening social responsibility as an effective constraint on individual behaviour.
73. The interaction between a diverse range of individual views and values, and the generally accepted truths and norms prevalent in society, inevitably involves challenges and disagreement. Content is not harmful just because it promotes an alternative view, theory, or ideology. Content is, however, harmful when it leads or

contributes to behaviour which causes harm. The diagram in the next section provides detailed descriptions of the different levels of harmful content.

74. Much harmful behaviour is illegal, and the explicit advocacy or incitement of the most serious examples is also illegal. In theory this should provide a powerful disincentive to the creation or publication of such content. In practice however, the dated nature of the relevant laws, uncertainty about their application to online content, the anonymous nature of much user-generated content and other enforcement challenges, significantly weaken their effectiveness as a deterrent. In addition, content can contribute to the development of harmful attitudes and behaviours, including illegal behaviours, in ways that are too subtle and indirect to be subject to legal sanction.
75. Not only do commercial incentives drive the creation and availability of harmful content, they also drive its active promotion on social media. There is clear evidence that people engage with content that triggers strong reactions including negative reactions. This means there is a huge commercial incentive for platforms to promote that kind of content to maximise consumer engagement. Promotion based on individual consumers' interaction histories can occur through algorithmic recommendations of:
 - similar content;
 - selective targeting of messages and advertisements; and
 - the prioritisation of "high impact" results on search engines.
76. This can lead to the duplication and repetition of the harmful characteristics of content and the amplification and reinforcement of the resulting harm.

Consumers have low incentives to avoid content that may lead them to harm others or society

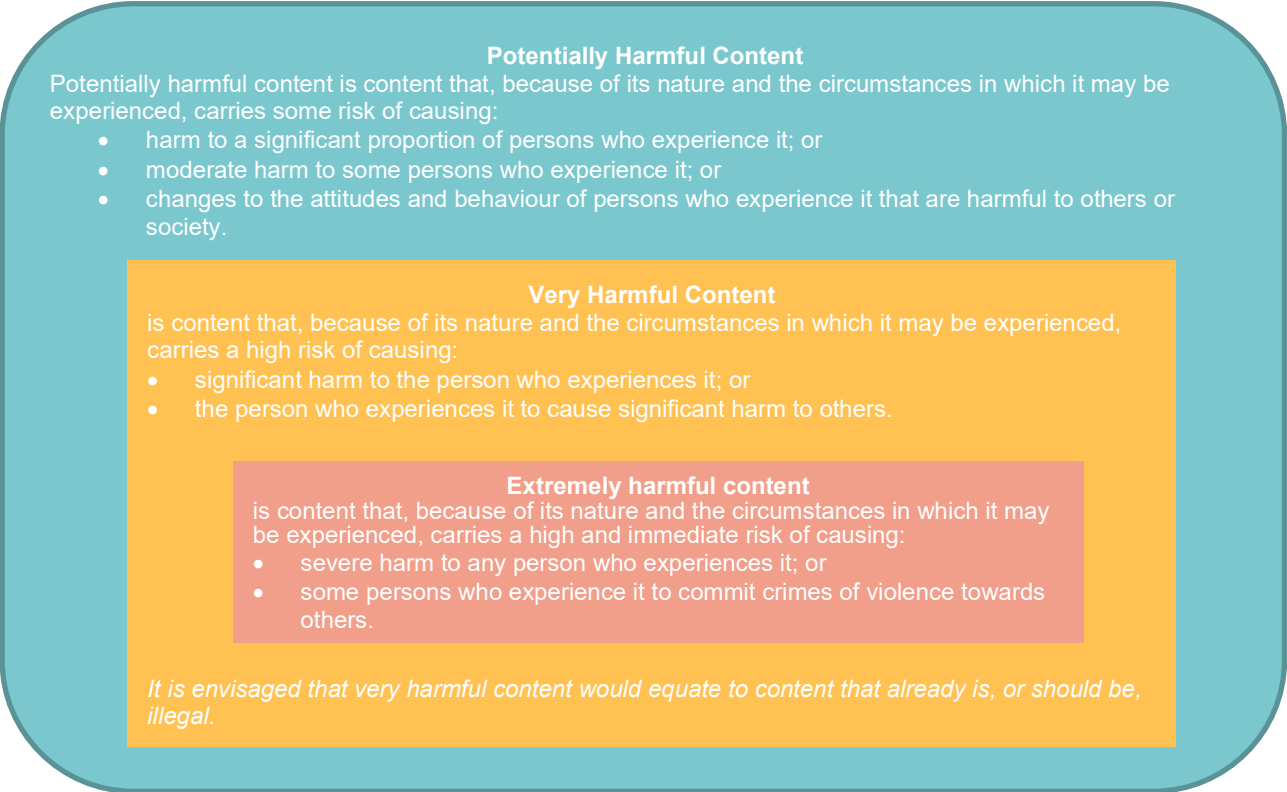
77. Some of the greatest risks of harm from interaction with content relate to the potential of some content to influence the behaviour of consumers in ways that result in harm to other individuals, to particular communities, or to society itself. It is important to emphasise that attitudes such as racism and sexism, extremist ideologies, and distrust of government that are reflected in this category of content are not within the scope of this analysis. The analysis is, however, concerned with the content itself and the impact it can have in leading to harmful behaviour.
78. As noted above, consumers typically have little understanding of how content can lead to harm and, in most cases, no information or warning about the potential impact of specific material. Improvements in these areas will enable them to avoid content where they wish to. However, incentives to avoid harming others and society are likely to be much weaker than self-protection and dependent on highly subjective matters relating to their world view, personal morality and understanding of harm. Those same views and beliefs may themselves constitute a strong motivation to seek and experience media content that conveys and reinforces them, overriding or obscuring any balanced assessment of the likelihood of harm.

What objectives are sought in relation to the policy problem?

There are nine objectives that are key to create an effective harm minimisation framework for regulating content in New Zealand

79. The overarching objective of the Review is to design a modern, flexible, and coherent regulatory framework to provide people with greater protection from harm that can be caused by content – irrespective of the way that content is delivered.
80. This overarching objective can be broken down into nine subsidiary objectives that either target consumers, creators (including users who create social media content), or publishers/providers. The nine objectives aim to either minimise the potentially harmful impacts of content, or disincentivise the creation and dissemination of harmful content.

- 81. Together, these nine objectives address the overarching problem of New Zealanders experiencing a wide range of unacceptable harms from media and online content that existing regulatory mechanisms are unable to address or mitigate.
- 82. The scope of the objectives is framed in terms of potentially harmful, very harmful or extremely harmful content which have the following relationship (the diagram below is not reflective of proportions of content in each category):



Objective 1 – Consumers are better able to identify and avoid or manage potentially harmful content

- 83. This objective is about the ability of consumers to avoid or manage harmful impacts of content on themselves or others. It relates to both consumers themselves, and to how they experience content. It reflects an assumption that people will be motivated to avoid or manage their own interactions with content if they know it may be harmful to them.

Objective 2 – Consumers are incentivised to avoid potentially harmful content

- 84. This objective is about the motivation of consumers to avoid content that has potentially harmful impacts on others, but may not be harmful to the person directly consuming that content. In other words, it addresses the harm that can occur because neither provider nor consumer has a direct interest in avoiding harm to others or society (e.g., the consumption of racist content leading to discriminatory attitudes towards the specific community that the content is about).

Objective 3 – Creators of content (including social media users) are better able to anticipate and avoid or minimise the potentially harmful impact of that content

- 85. This objective is about the ability (the understanding and knowledge) of persons to recognise how the content they create may cause harm, and how they can avoid or minimise the impact of that harm.

Objective 4 – Creators of content (including social media users) are incentivised to minimise the potentially harmful impact of that content

86. This objective seeks to motivate creators of content to minimise the potentially harmful impact of their content on consumers or others, given that creators will not be the ones who experience the potentially harmful impacts of their own content.

Objective 5 – Creators of content (including social media users) are incentivised not to create very harmful content

87. This objective seeks incentives for creators to avoid creating very harmful content that are sufficient to outweigh commercial and other incentives to create such content.

Objective 6 – Publishers/providers of media and online content are mandated to minimise the prevalence and impact of potentially harmful content

88. This objective seeks clear authority and obligations for MSPs to take action to reduce potential harm from content.

Objective 7 – Publishers/providers of media and online content are incentivised not to promote potentially harmful content

89. This objective seeks incentives for publishers/providers to avoid promoting potentially harmful content (e.g., by suggestions, recommendations, search result prioritisation etc) that are sufficient to outweigh commercial and other incentives to promote such content.

Objective 8 – Publishers/providers of media and online content are incentivised not to publish/provide very harmful content

90. This objective seeks incentives to avoid publishing/providing very harmful content that are sufficient to outweigh commercial and other incentives to publish/provide such *content*.

Objective 9 – Consumers are prevented from experiencing extremely harmful content

91. This objective seeks to avoid the immediate and severe harm that can be caused by experiencing the most extreme content. This includes content such as child sexual abuse material (CSAM) and violent extremist or terrorist material.

Section 2: Deciding upon an option to address the policy problem

What criteria will be used to compare options to the status quo?

92. There are options for each of the nine objectives that fall under either the Supportive, Balanced, or Prescriptive co-regulatory approaches. However, the same criteria will be used for all nine objectives, to compare the associated options to continuation of the status quo. Criteria 1-3 are weighted more than the other criteria, which are described in further detail below.

Criterion 1: Extent to which the objective is achieved compared to the status quo

93. This criterion compares the extent that the option is able to meet the specific objective, compared to maintaining the status quo. The option could be either significantly better, better, similar, worse or significantly worse than the status quo. The result will be given a weighting of three as it is the most important criterion in determining the potential benefit of the option in relation to the objective.

Criterion 2: Adaptability

94. This criterion examines the flexibility of the option, for example, how well the option can adapt to different media and online content types, future changes in the media and online environment, technological advances and changing societal expectations. This will also require evaluation against the flexibility that is needed to align with social and cultural values and the New Zealand context. This will be given a weighting of two as future proofing will be a crucial characteristic of the new framework.

Criterion 3: Proportionality

95. This criterion examines whether the expected benefits of the option justify its anticipated costs and other negative impacts (which could include, for example, if the option results in restriction of access to content that is not harmful or poses low risk of harm). This will be given a weighting of two to reflect the importance of options balancing the risks of harm they are seeking to address with the potential negative impacts of the option. Potential negative impacts may include unjustified limitation on freedom of expression, high costs of compliance for platforms and users, or enforcement difficulties for Government.

Criterion 4: Maximising benefits of content

96. This criterion examines the extent to which the option maintains the private and public benefits of media and online content, for example, the benefits of a free and open internet, existing standards of freedom of expression (including freedom of speech, freedom to access and provide information) guaranteed under the New Zealand Bill of Rights Act, economic growth, fostering social cohesion and sense of community/belonging, and entertainment. This does not have additional weighting because all feasible options would need to meet the existing threshold of being a justifiable limitation on the right to freedom of expression in a free and democratic society (as per Section 5 of the Bill of Rights Act).

Criterion 5: International Alignment

97. This criterion examines alignment with what comparable democratic jurisdictions²³ are doing in terms of supporting shared objectives, enabling joint action, and enabling

²³ Comparable jurisdictions include democratic states or groups of states such as Australia, United Kingdom, Ireland, Canada and the European Union.

consistency of expectations of global platforms. A summary of recent and current initiatives in those and some other jurisdictions is attached as **Appendix B**. This criterion does not have any additional weighting to avoid double counting where international alignment contributes materially to the effectiveness of an option (criterion 1). This criterion recognises that while other potential benefits of international alignment may be relevant to the relative merits of the proposed options, they do not in themselves affect the extent to which the options can achieve the objectives.

Criterion 6: Alignment with international rights-enhancing commitments

98. This criterion examines whether the options align with international instruments, treaties, and agreements that New Zealand has led, signed, or expressed support for that promote or enhance rights for everyone or the rights of particular groups of people.

Examples of international agreements include the Christchurch Call²⁴ and the Declaration for the Future of the Internet²⁵.

99. This also includes consideration of core international human rights instruments such as the:

- International Convention on the Elimination of All Forms of Racial Discrimination (ICERD);
- United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP);
- International Covenant on Civil and Political Rights (ICCPR);
- International Covenant on Economic, Social and Cultural Rights (ICESR);
- Convention on the Elimination of All Forms of Discrimination against Women (CEDAW);
- Convention on the Rights of the Child (CRC); and
- Convention on the Rights of Persons with Disabilities (CRPD).

100. Like criterion 5, this criterion does not have any additional weighting as a lack of international alignment with international commitments will not prevent the proposed option from achieving the objective. However, it is still relevant to the relative merits of the options.

What scope will options be considered within?

101. All options must be consistent with Te Tiriti o Waitangi, as well as with the framework of constitutional rights and freedoms embodied in the New Zealand Bill of Rights Act 1990, Human Rights Act 1993, and Privacy Act 2020. At this stage, all options are considered to be consistent with New Zealand's constitutional settings, but a final determination will only be possible when detailed policy proposals have been developed. An initial assessment of consistency with Te Tiriti o Waitangi principles can be found in Section 3.

102. All options are variations of an industry co-regulatory model, which involves Government working with industry to achieve better outcomes for content consumers,

²⁴ A commitment by Governments and tech companies to eliminate terrorist and violent extremist content online. <https://www.christchurchcall.com/call.html>

²⁵ The Declaration represents a political commitment from the United States and more than 60 global partners to advance a positive vision for the Internet and digital technologies, including an Internet that is truly open and fosters competition, privacy, and respect for human rights. <https://www.state.gov/declaration-for-the-future-of-the-internet>

creators, publishers and providers. The industry co-regulatory model is preferred and is underpinned by some assumptions:

- That there is role for Government stewardship and oversight in the regulation and moderation of media and online content, to reduce harms from this type of material, reflecting feedback from targeted engagement that government has a role to play in holding Media Service Providers more accountable;
- Direct government regulation is not preferred given the important freedom of expression rights engaged, and regulatory oversight is best achieved by an independent Regulator, operating at arms' length from the Government;
- Industry self-regulation is insufficient in addressing identified harms and would lead to further fragmentation of the system;
- Regulatory obligations emanating from a legislative framework (delivered through industry codes)²⁶ are necessary to ensure effective compliance and enforcement;
- Industry is best placed to moderate content for safety and to reduce the risk of harm on a daily basis, building on their expertise and knowledge; and
- Māori as Treaty partners, communities and civil society have a role to play in ensuring processes and ways to reduce harm reflect societal expectations.

103. The options are also set out so that responses to potentially harmful content apply to, but do not need to focus on, very harmful and extremely harmful content. This is because very harmful content will be responded to specifically in accordance with the options set out for **objective 5** and **objective 8**. Extremely harmful content will be responded to specifically in accordance with options under **objective 9**.

104. The options are underpinned by a number of other assumptions. These are:

- that people will be motivated to avoid or manage interaction with content if they know it may be harmful to them;
- that MSPs' tolerance or promotion of potentially harmful content reflects commercial incentives around maximising audience engagement, enabling immediate posting of user-generated content and enabling users to post content without delay or scrutiny; and
- that all options will continue to reflect international best practice methods and an internationally recognised harm environment.

Under each objective, there are three options that are variations of the industry co-regulatory model

105. The options under each objective are aligned with three approaches for industry co-regulation – Supportive, Balanced, and Prescriptive. These three approaches differ in how strongly Government might be involved within the co-regulatory model. This includes the degree to which Government is involved in regulating content, the strength of government expectation-setting, and the role and responsibility of Media Service Providers to minimise content-related harm.

²⁶ Refer pgs. 15 – 17 of the draft discussion document *A proposed new framework for regulating media and online content in Aotearoa New Zealand*

106. As the variations are primarily in the strength of Government's role, the levers under each approach below focus on the role for Government in the broad categories of:
- education and awareness;
 - moderation and deterrence; and
 - censorship.
107. The levers represent a mix of existing and proposed regulatory tools and is discussed under each approach below. These levers were tested with a variety of stakeholders as part of the targeted engagement process, aimed at helping the Department refine proposals for the new regulatory framework. However, the levers falling under each option are not necessarily mutually exclusive and there is possibility for some levers to intersect into multiple options.
108. Stakeholder feedback on the levers that could be used in the new framework are set out below:
- **Education and awareness:**
 - i. There was general agreement and support for education as a lever that will work for some of society.
 - ii. Many stakeholders echoed the point that the people who will engage with education efforts are the ones that need it least and questioned what could address those that are unlikely to engage with this. The framework is articulated best when it emphasises the role of each lever in addressing gaps, particularly where another lever might be limited.
 - iii. Feedback recommended carefully considering which groups we are trying to target through education. References also made on the idea that building critical thinking is not enough – this needs to be coupled with education in other areas.
 - **Moderation and deterrence:**
 - i. Enforcement agency stakeholders were supportive of having regulatory tools to support moderation.
 - ii. Stakeholders suggested the possibility of government agencies intervening to provide help, rather than trying to restrict or ban. For example, providing support for people affected by eating disorders and the idea of tailored intervention to bring people to the right path.
 - iii. Participants also believed that if we were to allow platforms to come up with their own codes, we should set high level of standards that is required e.g., reporting functions, impacts on algorithms, complaints mechanisms. They believed these should not be up to the agencies and that government should set standards for platforms.
 - **Censorship:**
 - i. Enforcement agency participants recommended that tools and terminology be well articulated to ensure interoperability with international enforcement agencies i.e., the more similar our laws to theirs, the more likely they are to act on New Zealand's behalf.
 - ii. Participants with enforcement roles advocated for the need for take-down and filtering mandates strong enough to deal with rogue providers who won't abide by New Zealand law. It was felt that there needs to be a clear specification of this, such as with a mandated code like the UK Online Safety Bill, and that if there are repeated warnings and violations then

service disruption mechanisms can be put into action to see the removal of content.

- iii. Stakeholders were supportive of current censorship processes even though they are spread across various government agencies. Some suggested that it would be good to see an information feedback loop between service providers, regulators and complaints made by users prior to decision making.

Option A: Continuation of the status quo

109. The **status quo** largely entails a continuation of fragmented Government and industry regulatory settings that result in different approaches being taken to content on different platforms. Content on some media platforms, for example social media content, is not regulated by Government at all. Content on some media platforms is subject to content warning requirements (e.g., cinematic releases, content on some streaming services), and content that is broadcasted is subject to programme restrictions. Some media sectors may also have public complaints processes. Section 1 provides a more detailed outline of the current regulatory situation (status quo) and how this is expected to develop if no action is taken.

110. As the status quo involves fragmented, overlapping, or voluntary approaches to content, the levers outlined below include both Government responsibilities and industry initiatives that are outside of Government regulation. Particular levers under the status quo are:

- **Education and awareness:**

- i. Government undertakes some public education (e.g., the *Keep It Real Online* and *A Bit Suss* campaigns) to educate consumers and creators of content on the dangers of some online content and behaviours;
- ii. Crown Entities such as the Classification Office and other organisations like Netsafe carry out research and publish reports into content related harms, with the aim of educating consumers and professional bodies and preventing further exposure to material that may cause harm;

- **Moderation and deterrence:**

- i. Government requires consumer warnings for specific content on some types of media (examples include streaming services and cinematic releases), that is rated according to whether the content should be restricted to a particular age group and/or other categories of legitimate users;
- ii. Government requires complaints processes for some types of media;
- iii. Government uses regulatory monitoring and enforcement measures to ensure industry is compliant with laws relating to potentially, very, and extremely harmful content that is subject to Government regulation;
- iv. Government continues its co-regulatory approach for content produced by or for broadcast media, which will continue to be subject to standards developed under the Broadcasting Act;
- v. Some MSPs adopt or sign up to voluntary codes, which include their own complaints processes (e.g., the Media Council and Netsafe have voluntary codes that their industry members can choose to comply with);
- vi. Some social media platforms have adopted community guidelines or standards. For example, their terms and conditions may expect users to

follow certain guidelines when creating content, and platforms may provide a complaints process to deal with breaches;

- vii. Creators of traditional curated content (i.e., not user-generated) will continue to be subject to industry standards or voluntary codes;
- viii. Certain publishers and providers are subject to professional standards or codes of practice that set out allowable practices (e.g., rules applying to mainstream news journalists, members of the Advertising Standards Authority, signatories of Christchurch Call). These, and fear of reputational damage, may incentivise MSPs against promoting potentially harmful content on their services;

- **Censorship:**

- i. Government is able to undertake interim assessments of content as 'objectionable', which allows immediate responses to remove/block extremely harmful content where urgency applies;
- ii. Government uses take-down or removal measures for 'objectionable' content when it identifies it. It may seize physical copies of such material and conduct criminal prosecutions against those publishing or providing 'objectionable' content;
- iii. Government imposes criminal and civil penalties for the creation and possession of extremely harmful (i.e. 'objectionable') content;
- iv. As content on some media is not regulated at all, extremely harmful content on some social media platforms may not be detected and appropriately banned. However, Government does rely on informal relationships with online content providers to remove extremely harmful content, to varying degrees of success;
- v. Content blocking/filtering is used on a voluntary basis by ISPs in a limited capacity for CSAM material.

Option B: Supportive approach

111. The **supportive approach** takes a primary focus on government supporting and encouraging industry and community initiatives to achieve the objectives. Under this approach, most of Government's effort would go into supporting and encouraging industry and community initiatives: for example, funding education and awareness initiatives. Standards in matters such as acceptable content, moderation practices and complaints procedures would largely be supported through industry-led codes. There would be minimal government intervention using monitoring and enforcement powers, except for (as at present) illegal 'objectionable' material. Under this approach, Government would rely on existing regulatory and law enforcement powers (such as

take down notices for illegal content and voluntary filters) operating within existing legislative constraints.

112. Particular levers under the supportive approach could include:

- **Education and awareness:**

- i. Government provides or funds public awareness campaigns to inform consumers and creators of content about the potential harmful and divisive impacts of different types of content, including mis and disinformation;
- ii. Government potentially makes changes to the education curriculum to better educate young consumers about potentially harmful content;
- iii. Government encourages and enables community-led (non-industry) initiatives to educate consumers about risks surrounding potentially harmful content;
- iv. Government promotes content creators' participation in community-led education and awareness initiatives, so creators are educated on the potential risks of harm from content they create;
- v. Government promotes industry's (Media Service Providers or their representative bodies) participation in community-led (non-industry) initiatives, to educate themselves about the prevalence of potentially harmful content on their media and online services and impacts on their consumers.
- vi. Government encourages industry's (Media Service Providers or their representative bodies) to have information and education initiatives that support consumers to be aware of risks in accessing potentially harmful content, empower consumers to avoid particularly risky content, and manage the impact of harmful interaction with content;
- vii. Government facilitates and encourages information and education initiatives by industry (Media Service Providers or their representative bodies) that aim to enable creators to be aware of risks incumbent in creating potentially harmful content, very harmful content, and extremely harmful content;
- viii. Government encourages and incentivises consumer participation in community-led (non-industry) education initiatives on the negative impacts of extremely harmful content. This supports consumers to not attempt to access such content;
- ix. Government conducts its own educative and informational initiatives around the risks posed by extremely harmful content;

- **Moderation and deterrence:**

- i. Government encourages (through appropriate incentives or relationships), individual publishers or providers to join or participate in industry bodies, which foster a common focus on preventing the provision and publication of potentially harmful and very harmful content;
- ii. Government promotes industry to work together to create best practice codes of conduct and industry-wide standards and encourage this to be rolled out for content creators;
- iii. Government supports publishers and providers (industry) to adopt minimum standards for minimising harms from content, and supports and

encourages the upkeep of any harm minimisation codes that are developed by industry;

- iv. Government encourages publishers and providers (industry) to provide opportunities for access and safety controls on their services;
 - v. Government encourages (through appropriate incentives or existing relationships) industry adherence to professional standards or codes of conduct for not promoting harmful content;
 - vi. Government encourages industry quality assurance schemes to ensure potentially harmful content is appropriately captured and responded to in a timely manner. Support would be provided to publishers and providers that do not meet expectations;
 - vii. Government uses regulatory monitoring and enforcement measures to encourage industry to comply with Government laws and regulations relating to very harmful content;
 - viii. Government encourages industry adherence to codes of conduct or practice that prioritise not publishing or providing very harmful content. This includes Government administering (at arm's length) quality assurance schemes to support industry adherence to safe practices relating to not publishing very harmful content;
 - ix. Government encourages (through appropriate incentives) industry bodies' enforcement efforts against their members who may have provided or published very harmful content;
- **Censorship:**
 - i. Government encourages (through appropriate incentives) industry use of consumer warnings around extremely harmful content;
 - ii. Government encourages industry adherence to codes of conduct/standards that minimise prevalence of extremely harmful content on platforms.
 - iii. Government encourages (through appropriate incentives) industry complaint/alert processes relating to extremely harmful content. This supports effective removal of such content from platforms; and
 - iv. Government continues to remove or block access to content that is considered illegal in New Zealand (i.e., 'objectionable' material).

Option C: Balanced approach

113. The **balanced approach** is a mix of education, government regulatory intervention, and partnering with industry, Māori as Treaty partners, and communities to achieve the objectives. The expectations on MSPs are set out in legislation, which is likely to be principles-based and set out required processes for MSPs to meet certain outcomes. Much of the responsibility to reduce harm sits with the MSPs, using the processes required under the legislation. There would be some new offences and penalties in relation to enforcing compliance with codes, including greater powers to require take

down of material but the government's enforcement role relates mostly to how MSPs are implementing the required processes and dealing with illegal material.

114. Particular levers under the balanced approach could include:

- **Education and awareness:**

- i. Government provides or funds public awareness campaigns to inform consumers and creators of content about the potential harmful and divisive impacts of different types of content, including mis and disinformation;
- ii. Government potentially makes changes to the education curriculum to better educate young consumers about potentially harmful content;
- iii. Government encourages community-led (non-industry) groups to educate consumers about the risks attaching to potentially harmful content within the context of their communities;
- iv. Government facilitates and encourages industry information and education initiatives to enable content creators to be aware of potential risks of harm from content they create;
- v. Government promotes content creators' participation in community-led education and awareness initiatives, so creators are educated on the potential risks of harm from content they create;
- vi. Government facilitates and encourages information and education initiatives by industry (Media Service Providers or their representative bodies) on their services, that enable consumers to be aware of risks in accessing potentially harmful content, to increase resilience for the consumer when harmful content is viewed, and what to do when consumers come across harmful content on their services (e.g., guidance on complaints process);
- vii. Government works with community stakeholders to educate community members about the risks of creating very harmful content, and the legal consequences of doing so;
- viii. Government provides funding and works with community stakeholders to educate community members on how to avoid being harmed by extremely harmful content. Government may also conduct independent educative initiatives, raising community awareness of the dangers of extremely harmful content;

- **Moderation and deterrence:**

- i. Government requires publishers and providers (MSPs) to adopt harm minimisation codes that meet minimum standards in relation to reducing the prevalence and impact of potentially harmful content. This includes expectations in respect of consumer warnings, educating consumers about potentially harmful content and associated risks, and public complaints or alert processes;
- ii. Government requires MSPs to adopt harm minimisation codes that meet minimum standards in relation to the identification of potentially harmful content (and sites that host such content) and the de-prioritisation of such content/sites in recommendations, search results, etc. on their services;
- iii. Government ensures that regular assessment and reporting requirements are applied to publishers and providers regarding the prevalence of potentially harmful content. This enables the Government to hold publishers

and providers accountable for such content on their services and mandates a more proactive approach to not promoting such content;

- iv. Government encourages industry professional standards that may require consumer information above that required by Government in relation to potentially harmful content;
 - v. Government encourages industry enforcement efforts (sanctions/fines/notices/exclusion of benefits/complaints investigation) in the case of non-compliance with the agreed to industry-wide standards;
 - vi. Government imposes criminal or civil penalties for creation, possession, publication and provision of very harmful content. This measure reduces prevalence of such content on platforms, reducing likelihood that consumers experience such content;
 - vii. Government creates a new offence of promoting very harmful content, with civil or criminal penalties;
 - viii. Government uses regulatory monitoring and enforcement measures to ensure industry is compliant with Government laws and regulations relating to very harmful content;
 - ix. Government produces consumer warnings in relation to extremely harmful content, and disseminates them through government channels, to educate consumers against accessing such content;
 - x. Government encourages industry's use of effective complaint/alert processes, to be used by consumers in the event of discovery of extremely harmful content, and to support industry removal of such content;
- **Censorship:**
 - i. Government uses take-down and removal of content measures to incentivise creators to not create very harmful content;
 - ii. Government takes steps to ensure that publishers and providers who repeatedly promote very harmful content are not able to participate in the New Zealand market;
 - iii. Government uses regulatory monitoring and enforcement measures to ensure industry is compliant with Government laws and regulations relating to extremely harmful content, and that such content is not found on platforms. This includes use of take-down and removal of that content. Non-compliance with these requirements may lead to high penalties, or exclusion from the market.
 - iv. Government continues to remove or block access to content that is considered illegal in New Zealand (i.e., 'objectionable' material).

Option D: Prescriptive approach

115. The **prescriptive approach** primarily focuses on government regulatory intervention to achieve the objectives, with less emphasis on education and voluntary uptake of harm minimisation approaches. The legislation would be much more directive to industry, with detailed outcome requirements and rules for meeting those requirements and could include powers for tools such as a mandatory filter for illegal content. The regulator takes on a bigger role in monitoring, enforcing, and prosecuting regulatory

non-compliance and criminal breaches of the law. Particular levers under the prescriptive approach could include:

- **Education and awareness:**
 - i. Government could provide its own consumer warnings to consumers, independent to industry, to ensure consumers are well educated on risks of potentially harmful content.
- **Moderation and deterrence:**
 - i. Government mandates and monitors industry compliance with prescribed harm minimisation codes in respect to consumer warnings, educating consumers about potentially harmful content and associated risks, and public complaints or alert processes. These could include requirements to monitor and mitigate repeated access to potentially harmful content;
 - ii. Government ensures that strict reporting requirements are applied to publishers and providers of content regarding the prevalence of potentially harmful content on their services, to hold them accountable for such content included where this is promoted on their services;
 - iii. Government prescribes harm risk thresholds (lower than the threshold for very harmful content) that must not be exceeded by content promoted or recommended by media platforms or publishers, with criminal and civil sanctions for exceeding them. This includes maintaining a register of sites that must not be promoted;
 - iv. Government prescribes minimum standards that content creators must comply with in relation to potentially harmful content. Repeated breaches of minimum standards, by creators, publishers or providers is subject to civil or criminal liability;
 - v. Government requires industry to make efforts to prevent content creators from “publishing” potentially harmful content;
 - vi. Government uses civil or criminal penalties to ensure MSP (industry) compliance with harm minimisation requirements;
 - vii. Government criminalises the publication and provision of very harmful content. This measure reduces prevalence of such content on platforms, reducing likelihood that consumers access or experience such content;
- **Censorship:**
 - i. Government uses regulatory monitoring and enforcement measures to ensure industry is compliant with Government laws and regulations relating to very harmful and extremely harmful content, and that such content is not found on platforms. This includes greater use of take-down and removal of

offending content. Government may also explore the use of filtering/blocking mechanisms for extremely harmful content;

- ii. Government may use “service disruptions”²⁷ as a last resort measure for persistent and repeated breaches of minimum standards by content creators;
- iii. Government prohibits the promotion of content that exceeds prescribed harm risk thresholds and of sites listed on a formal register;
- iv. Government reserves the power to ensure that publishers or providers services’, on which potentially harmful content is prevalent, are not able to participate in the New Zealand market;
- v. Government has exclusion powers to control the access to the New Zealand market of media providers or publishers that are more likely to publish extremely harmful content;
- vi. Government continues to remove or block access to content that is considered illegal in New Zealand (i.e., ‘objectionable’ material).

Analysis of options

116. The four options described in this section (Status quo, Supportive approach, Balanced approach and Prescriptive approach) will be assessed against each of the 9 objectives laid out from sections 2.1 to 2.9. For each objective, the four options will be assessed by the 6 Criteria:

- Criterion 1: Extent to which the objective is achieved compared to the status quo (weighted by 3)
- Criterion 2: Adaptability (weighted by 2)
- Criterion 3: Proportionality (weighted by 2)
- Criterion 4: Maximising benefits of content
- Criterion 5: International Alignment
- Criterion 6: Alignment with international

²⁷ This refers to the ability to allow for regulators to take action to disrupt services of an MSP. This is done through applying to Courts for an order to restrict other services that facilitate the provision of the MSP’s services, including but not limited to domain name systems, online search services, hosting providers and advertising services.

Section 2.1: Objective 1 - Consumers are better able to identify and avoid or manage potentially harmful content

What options are being considered?

117. This objective is about the ability of consumers to avoid or manage harmful impacts on themselves or others. It relates to both consumers themselves, and how to they experience content. It reflects an assumption that people will be motivated to avoid or manage their own interactions with content if they know it may be harmful to them.

Option 1A – Status Quo

Analysis

118. Under a continuation of the status quo, this objective is unlikely to be met because different approaches to supporting consumers to avoid or manage harmful content across different media channels and existing gaps would continue. These inconsistencies mean the status quo is unlikely to meet the objective to improve the ability of consumers to avoid or manage harmful impacts on themselves or others.
119. Current Government regulation lacks agility because it is governed by prescriptive statutes and regulations. The regulation of broadcasted content has more flexibility as it relies on standards developed in conjunction with industry to give effect to statutory requirements, and these can be periodically reviewed. Outside of Government regulation, there are approaches that do undergo regular review to reflect and adapt to contemporary operating environments, but these approaches operate on a voluntary basis. Existing Government responses are proportionate to the risks on the media channels that are currently regulated. This excludes most online content, which largely relies on industry bodies and MSPs to provide consumer warnings and mitigate harmful content. This means that the ways harmful content is addressed by different MSPs will likely be inconsistent and may not always be proportionate to the risk of harm.
120. The status quo does little to preserve or enhance the private and public benefits of content. While existing regulation imposes few (and only justified) limitations on freedoms of speech and access, and only on some platforms, the fragmented and incomplete coverage does not enhance trust in the safety of the internet or preserve the freedoms of those who are intimidated or harmed by being targets of harmful content.
121. The status quo is also inconsistent with developments in comparable jurisdictions, where Governments are extending content regulation to the online content environment, and this function is generally undertaken by a single expert agency. The status quo does not prevent New Zealand from meeting its obligations under international instruments and treaties but provides limited mechanisms to enable this.

Option 1B – Supportive approach

Analysis

122. Under a supportive approach, the objective would be better achieved than under the status quo. This is because this option would involve Government funding and encouraging education programmes to educate consumers about potentially harmful content. The key focus would be on education and working with communities, and encouraging industry to provide information and education on potentially harmful content and risks to consumers.
123. This option would be significantly better than the status quo for adaptability as the approach will not be specific to media channel types. A multi-levered approach would be built into the regulatory framework, which allows for flexibility in mitigation and response efforts. In addition, a focus on industry and community-led initiatives (with

Government support) would enable the system to be flexible to respond to changes in the nature of harms and societal expectations.

124. The proportionality of this approach to the risk of harm would be better than the status quo, as the Government would support an industry and community-led approach to educating consumers about potentially harmful content, including online content. Through a strong focus on education initiatives, consumers would be supported to feel safer and more confident when using the internet. In turn, they would be better able to access the benefits of the online content environment.
125. The focus on supporting industry and community initiatives reflects an education and encouragement approach that many comparable jurisdictions have taken to harmful content, particularly in the area of mis/disinformation. This option will also provide avenues for New Zealand to meet some commitments under the Christchurch Call and core human rights instruments, but this is limited by the lack of enforceable mandates.

Option 1C – Balanced approach

Analysis

126. Under a balanced approach, the objective would be achieved somewhat better than the status quo. The option enables the Government to mandate publishers/providers to adopt codes that meet minimum standards in respect of consumer warnings, and public complaint/alert processes. This would build on and align existing initiatives by platforms and industries in respect of these tools, creating consistent standards for signalling harmful content, and improving the ability of consumers to identify and avoid/manage potentially harmful content. In addition, this option also involves government funding and encouragement for education programmes to make consumers better informed about identifying potentially harmful content.
127. This approach would be better than the status quo for adaptability, as it will apply across all media channel types. The multi-levered approach would allow flexibility in mitigation and response efforts. However, reliance on Government to incentivise harm minimisation initiatives may hinder adaptability and rapid responses to emerging issues as these initiatives can take time to fully implement.
128. This would allow much greater proportionality than the status quo as Government and industry could take a more proactive approach to responding to potentially harmful content utilising a range of levers that allow proportionate responses depending on the risk of harm.
129. Consumers would be enabled to feel safer and have greater confidence in using the internet, allowing better access to the benefits of the internet. The option would also be much better than the status quo in relation to international alignment, in terms of requiring industry led responses within a prescribed regulatory framework, with accompanying Government support for education and future-focused initiatives. This option also provides more avenues for New Zealand to meet its commitments under the Christchurch Call to address terrorist and violent extremist content, as well as the United Nations' core international human rights instruments.

Option 1D – Prescriptive approach

Analysis

130. Under the prescriptive approach, the objective would be achieved somewhat better than under the status quo. The option would allow Government to prescribe consumer warning requirements, harm mitigation measures and public complaint/alert processes that publishers and providers of content must meet. Much like a balanced approach there would be better consistency of content warnings, however, under this approach Government may provide its own consumer warnings, independent from industry. Civil

or criminal liability for non-compliance would ensure industry bodies meet obligations to ensure consumers are better able to identify potentially harmful content.

131. This option would be less adaptable than the status quo as legislated requirements, which are time-consuming to change, would replace voluntary industry-led codes which are generally better able to adapt quickly to changing harms.
132. This option would be more proportionate than the status quo, in that the Government would take a proactive approach to responding to potentially harmful content across all platforms. This includes mandating measures that address harm from content and monitoring compliance. However, legislated minimum standards across all platforms industry may not be proportionate to the risk of harm in all circumstances.
133. This option would be worse than the status quo in maximising benefits of content, as it relies largely on regulatory sanctions to ensure adherence to regulatory standards for addressing potentially harmful content. This may create incentives for MSPs to be overly cautious in moderating or removing content. This can have implications on loss of access for New Zealanders to content that would otherwise have been allowed.
134. This option is better than the status quo in relation to New Zealand's international alignment as there is a growing realisation in other jurisdictions that strong intervention mechanisms are required to address a diversifying array of potentially harmful content. The enforceable mandates under this option would also provide stronger avenues for New Zealand to meet its commitments under the Christchurch Call and other core human rights instruments.

How do the options compare to the status quo/counterfactual?

Objective 1: consumers are better able to identify and avoid or manage potentially harmful content

	Option 1A – Status Quo	Option 1B – Supportive approach	Option 1C – Balanced approach	Option 1D– Prescriptive approach
Extent to which objective is achieved (x3) To what extent is the option able to meet the objective? Partially/fully/not at all	0 Fragmented Government regulatory regimes mean that uneven approaches are taken to supporting consumers to avoid or manage harmful content.	+	+	+
Adaptability (x2) How flexible is the option? For example, how well can this option adapt to different media and content types, future changes in the media landscape, technological advances, changing societal expectations? Evaluate flexibility that's needed to align with cultural values and NZ context	0 In the areas where content is subject to Government regulation, adaptability is limited and constrained by media channel. There are voluntary approaches which have the ability to be flexible.	++	+	-
Proportionality (x2) Do the expected benefits of the option justify its costs/negative impacts? e.g. would the option result in restriction of access to content that isn't harmful?	0 Existing Government responses are proportionate to the particular media channels that are currently regulated. Inconsistent approach to ad hoc voluntary initiatives for content outside of current regulations.	+	++	+
Maximising benefits of content (x1) To what extent does the option preserve the private and public benefits of content?	0 Status quo maximises freedom of speech and access on platforms that are not currently regulated by Government, however, this also inhibits freedoms of people or groups who experience harm arising from the unregulated nature of these platforms.	+	+	-
International Alignment (x1) Does the option align with what other comparable jurisdictions are doing?	0 Several disparities to regulation compared to other jurisdictions but emphasis on freedom of expression and acknowledgement that users are experiencing harm is consistent.	+	++	+
Alignment with International Commitments (x1) Does the option align with international instruments, treaties, agreements, etc that New Zealand has led, signed on to or expressed support for?	0 The status quo does not disable New Zealand to meet its obligations under various instruments and treaties. Existing accessibility requirements are patchy.	0	+	++
Overall Assessment	0	11	13	5

Section 2.2: Objective 2 – Consumers are incentivised to avoid potentially harmful content

What options are being considered?

135. This objective is about the motivation of consumers to avoid content that has potentially harmful impacts on others, but may not be harmful to the person consuming that content. In other words, it addresses the harm that can occur because neither provider nor consumer has a direct interest in avoiding harm to others or society (e.g., the consumption of racist content leading to discriminatory attitudes towards the specific community that the content is about).

Option 2A – Status Quo

Analysis

136. The status quo provides fragmented and uncoordinated initiatives that have some success in raising awareness of the harmful impact of some content, but the focus is largely on reducing harm to consumers rather than to others. Most initiatives are ad hoc which means they are generally relatively easy to modernise or replace, but the fragmented, incomplete and uncoordinated approach will limit their effectiveness. Ad hoc initiatives may not always be proportionate to the risk of harm. While freedom of access is not limited, especially on platforms that are not currently regulated by Government, status quo initiatives do not explicitly focus on impacts on social cohesion, which can detract from freedom of access for target groups.

137. The status quo is similar to many international jurisdictions, where various agencies will provide initiatives for consumers, but these are generally more advanced and accessible in comparison. While New Zealand's ability to meet its obligations under various international treaties and instruments is not hindered under the status quo, it generally allows relatively weak responses.

Option 2B – Supportive approach

Analysis

138. This approach would be slightly better than the status quo at achieving the objective. The focus on improving consumer awareness of harm risk through consumer warnings and education initiatives will support motivation to avoid self-harm. However, this approach may have limited effects in motivating consumers to avoid content that may result in harm to others/society.

139. This approach would be more adaptable than the status quo because as the approach would be generic and not limited to specific media channels. Reliance on community and industry initiatives would allow greater responsiveness than relying solely on Government measures.

140. Modest improvements in consumer safety and awareness may also improve the ability and willingness of media to engage with initiatives, which in turn may improve perceived safety and trust to realise access benefits. This approach is also better than the status quo for international alignment and meeting international commitments. While the approach is largely the same as the status quo, the focus on coordinated education initiatives appears to be a relatively successful approach in international jurisdictions for incentivising consumers to avoid harmful potentially harmful content.

Option 2C – Balanced approach

Analysis

141. This approach would be more effective than the status quo in meeting the objective. Requiring MSPs to adopt harm minimisation codes that meet minimum standards for

consumer warnings and complaint/remediation processes will achieve a more certain, coordinated and consistent outcome across all media channels. Coupled with generic education initiatives, this should increase consumer motivation to avoid harmful content generally. However, greater consumer awareness may have only limited effects in motivating avoidance of content that may result in harm to others/society.

142. This would be better than the status quo for adaptability because the approach will be comprehensive and not specific to any media-channels. There would be some limitations in terms of adaptability and responsiveness due to partial reliance on Government mandates and funding. This option could be more proportionate than the status quo for proportionality as, within the mandated codes, MSPs could employ a range of levers appropriate to their content and channel.
143. As with the supportive approach, modest improvements in consumer safety and awareness may increase trust in the safety of engaging with media and allow more people to feel free to access content. This approach is also better than the status quo for international alignment as mandated platform codes are common levers that have been successful in international jurisdictions in educating consumers about harmful content overseas.
144. This would also be much better than the status quo for alignment with international commitments as the range of levers would enable Government and regulators to ensure groups at higher risk of harm (such as minority groups) are better protected from harmful content.

Option 2D – Prescriptive approach

Analysis

145. This approach would be more effective than the status quo to meet the objective. The levers under this option would incentivise MSPs to support consumers to avoid potentially harmful content through Government mandates. Obligations on both MSPs and consumers would be generic across all media channels.
146. The reliance on Government mandates would limit adaptability and delay responsiveness to emerging harms and changes in circumstances, offering no advantages over the status quo in this respect.
147. Stronger Government mandates and more rigid requirements increase the risk of requirements on providers that are disproportionate to risks, making this option worse than the status quo in terms of proportionality.
148. As with other options, moderate improvements in consumer safety would encourage greater access to the benefits of accessing content. This is better than the status quo.
149. This option would be worse than the status quo in terms of international alignment, as reliance on regulatory sanctions in other jurisdictions has had adverse impacts on internationally valued objectives in comparable jurisdictions, including freedom of expression and a free and open internet. However, this option is better than the status quo for aligning with international commitments by allowing for additional levers such as Government mandates to ensure groups that are at risk of harm are better protected in relation to content.

How do the options compare to the status quo/counterfactual?

Objective 2: Consumers are incentivised to avoid potentially harmful content

	Option 2A – Status Quo	Option 2B – Supportive approach	Option 2C – Balanced approach	Option 2D – Prescriptive approach
Extent to which objective is achieved (x3) To what extent is the option able to meet the objective? Partially/fully/not at all	0 Fragmented and uncoordinated initiatives have some success in raising awareness of some harms, but largely in relation to potential harms to consumers Status quo does not explicitly focus on initiatives to foster social cohesion.	+	+	+
Adaptability (x2) How flexible is the option? For example, how well can this option adapt to different media and content types, future changes in the media landscape, technological advances, changing societal expectations? Evaluate flexibility that's needed to align with cultural values and NZ context	0 Most initiatives are relatively ad hoc so can be modernised or replaced relatively easily, but fragmented and uncoordinated approaches limit will limit effectiveness.	++	+	0 No better or worse than the status quo in that approaches will be comprehensive and non-media specific, but worse in terms of limited adaptability and delayed responsiveness because of reliance on Government mandates.
Proportionality (x2) Do the expected benefits of the option justify its costs/negative impacts? e.g. would the option result in restriction of access to content that isn't harmful?	0 Limited proportionality from current consumer awareness/information focussed measures.	+	++	- Stronger mandates and more rigid requirements increase risk of regulatory overreach in terms of disproportionate requirements on providers with limited resourcing.
Maximising benefits of content (x1) To what extent does the option preserve the private and public benefits of content? Weighting: x1	0 Status quo maximises freedom of access, especially on platforms that are not currently regulated by Government. Status quo does not explicitly focus on initiatives to foster social cohesion.	+	+	0 Modest improvements in consumer safety and awareness may also improve ability/willingness to engage with media.
International Alignment (x1) Does the option align with what other comparable jurisdictions are doing?	0 Reflects that of many international jurisdictions, where consumer education and incentivisation initiatives are often executed by a variation of agencies. However, such education and incentivization efforts are frequently more advanced and accessible than those of New Zealand.	+	+	- Worse as regulatory sanctions to incentivise consumers have resulted in over-regulation of potentially harmful content in other jurisdictions, leading to adverse impacts on freedom of expression and a free internet.
Alignment with International Commitments (x1) Does the option align with international instruments, treaties, agreements, etc that New Zealand has led, signed on to or expressed support for?	0 The status quo does not hinder New Zealand to meet its obligations under various instruments and treaties.	+	++	0 This option does not hinder New Zealand to meet its obligations under various instruments and treaties. Better as all media channels would be captured.
Overall Assessment	0	12	13	2

Section 2.3: Objective 3- Creators of content (including social media users) are better able to anticipate and avoid or minimise the potentially harmful impact of that content

What options are being considered?

150. This objective is about the ability (the understanding and knowledge) of persons to recognise how the content they create may cause harm, and how they can avoid or minimise the impact of that harm.

Option 3A – Status Quo

Analysis

151. The status quo provides ad hoc Government education initiatives for creators of user-generated content to understand the impacts of, and techniques to avoid, harmful content. In addition, creators of traditional content (not user-generated) continue to be subject to regulations and voluntary codes that are dependent on which media channels the content they create is made available on.
152. The status quo also has existing but fragmented, Government regulatory monitoring and compliance functions. There is limited adaptability for existing Government regulation as it is constrained by media channel types. The responses to harm occur as they arise which means it is often a proportional response.
153. The status quo has several disparities to comparable international jurisdictions, where content regulation is enforced by a single expert agency. However, the status quo does not hinder New Zealand to meet its obligations under various international instruments and treaties.

Option 3B – Supportive approach

Analysis

154. This approach would be better than the status quo in achieving this objective as the Government would make a concerted effort to support, and possibly fund, industry and community initiatives aimed at educating creators on how content can cause harm. This will also be consistent across all media channels.
155. Adaptability would be much better than the status quo because this approach focuses on industry and community-led initiatives with Government playing a supporting role. This allows for flexibility to adapt to the changing nature of harms and societal expectations, so that creators can continue to anticipate and avoid/minimise the harmful impact of content. This option would largely be the same as the status quo for proportionality, but with more coordinated industry and education sector initiatives.
156. On balance, this option remains the same as the status quo for benefits of content. While there may be higher harm minimisation expectations relating to content creation that may have an inhibiting effect on the creation of some content, this is justified and offset by enabling a safer environment for consumers.
157. This would be better than the status quo for international alignment and meeting international commitments. This is because it has a strong focus on education which is similar to the approach taken by many Scandinavian countries. There would also be a concerted effort to support and educate the public on harms associated with certain types of content, such as racist comments and discriminatory posts. This would support

New Zealand meeting commitments under the Christchurch Call and other core human rights instruments.

Option 3C – Balanced approach

Analysis

158. This approach would be better than the status quo in achieving this objective as there would be a coordinated and platform-neutral effort, encouraged by Government, for industry-led setting of codes of conduct and minimum standards for minimising harms from content. This may lead to more content that is created with regard for how content can cause harm. Government would also play a role in raising public awareness of how content causes harm, which may have a positive effect on the creation of less user-generated content that is harmful.
159. This would be better than the status quo for adaptability as the approach would be platform-neutral and capture creators across all media channels. However, there is a reliance on Government to incentivise or fund harm minimisation initiatives and on engaging communities which may slow the ability to adapt to emerging harms. This would be better than the status quo for proportionality as Government takes a more proactive approach to responding to potentially harmful content. This is done through a range of levers, including working with industry which allows for different risk profiles and circumstances of MSPs to be considered.
160. On balance this approach is not better or worse than the status quo for benefits of content. While there may be higher harm minimisation expectations relating to content creation that may impact others' freedom of speech, this is justified when considered against the benefits to freedom from discrimination and harm.
161. This option is better than the status quo in international alignment as it reflects the proactive, education focused approach taken by comparable jurisdictions. This is much better than the status quo for meeting international commitments as there would be a concerted effort to support and educate the public on harms through a range of different levers. This also provides more avenues for New Zealand to meet its commitments and the Christchurch Call along with other core human rights instruments.

Option 3D – Prescriptive approach

Analysis

162. This approach would be much better than the status quo as the Government would be able to mandate industry use of codes of conducts or practices to minimise risks from potentially harmful content, including the publication and widespread dissemination of such content. These codes would be consistent across providers as they would be aligned with standards set by the government. This would result in content creators being better able to understand what content can have harmful impacts and avoid creating such content. Content creators are further discouraged by the potential for criminal or civil liability for the creation of potentially harmful content.
163. This would be worse than the status quo for adaptability because there would be a heavy reliance on Government to incentivise or fund harm minimisation initiatives and industry would rely on Government to provide codes. This reliance on Government mandates may hinder adaptability and generate a slower response to emerging harms/issues. This would also be worse than the status quo for proportionality due to the approach relying on Government mandates. This creates a blanket approach to harms which may not be appropriate for certain types of content or MSPs.
164. This option relies on regulatory sanctions to ensure adherence to regulatory mechanisms applying to potentially harmful content. This is worse than the status quo for benefits of the media as it may create a perverse incentive for MSPs to be overly

cautious in moderating or removing content, leading to potential loss of access to content and creators of content avoiding creation of content that may be legitimate.

165. This approach would be more aligned than the status quo with e stronger regulatory measures adopted by some other comparable jurisdictions. However, reliance on regulatory sanctions to ensure adherence could risk regulatory overreach and unjustified limitations on access. This would make the approach worse than the status quo for meeting international commitments.

How do the options compare to the status quo/counterfactual?

Objective 3: Creators of content (including social media users) are better able to anticipate and avoid/minimise the harmful impact of the content

	Option 3A – Status Quo	Option 3B – Supportive approach	Option 3C – Balanced approach	Option 3D – Prescriptive approach
Extent to which objective is achieved To what extent is the option able to meet the objective? Partially/fully/not at all Weighting: x3	0 Education through campaigns and consumer warnings on potentially harmful content. Creators of traditional editorial content (i.e. not user-generated) are subject to broadcasting standards and/or the classification system and/or voluntary codes that are media channel specific. Fragmented, Government regulatory monitoring and compliance functions exist.	+ Partially achieves the objective through Government supporting and funding industry/community-led initiatives to educate creators on the impact of harmful content.	+ Partially achieves the objective through a coordinated and platform-neutral effort with encouragement from Government, for industry-led setting of codes and minimum standards. Government plays a role in raising public awareness about the impact of harmful content.	++ Achieves the objective by Government mandates for industry to use codes of conduct or practice to minimise risks from potentially harmful media content. There would be criminal or civil liability for the creation of potentially harmful content.
Adaptability How flexible is the option? For example, how well can this option adapt to different media and content types, future changes in the media landscape, technological advances, changing societal expectations? Evaluate flexibility that's needed to align with cultural values and NZ context Weighting: x2	0 In the areas where content is subject to Government regulation, adaptability is limited and are constrained by media channel. Voluntary codes exist but flexibility still constrained by the outdated overarching system in which they operate.	++ Highly adaptable through industry/community-led initiatives (with Government support) would enable flexibility, particularly to changing nature of harms and societal expectations.	+ Some adaptability through a range of levers that are not media channel specific. Flexibility slightly hindered by reliance on Government to incentivise or fund harm minimisation initiatives, and on engaging communities.	- Limited adaptability due to heavy reliance on Government to incentivise or fund harm minimisation initiatives, and on engaging communities.
Proportionality Do the expected benefits of the option justify its costs/negative impacts? e.g. would the option result in restriction of access to content that isn't harmful? Weighting: x2	0 As responses are targeted to specific issues as they arise, responses under the status quo are proportional to the specific harm identified.	0 This option would largely be a continuation of the status quo, but more coordinated across industry and education sector initiatives.	+ Proportional as Government takes a more proactive approach to responding to potentially harmful content, utilising a range of levers – some led by industry.	- Not always proportional due to focus on Government mandates which takes a blanket approach that may not be appropriate for certain types of content or MSPs.
Maximising benefits of content To what extent does the option preserve the private and public benefits of content? Weighting: x1	0 Status quo maximises content creators' freedom of expression, and user access to their content, on platforms that are not currently regulated by Government. However, this also impacts on the freedoms of people or groups who experience harm.	0 This is largely a continuation of the status quo, with slightly higher harm minimisation expectations from creators of content.	0 On balance, this option doesn't add or take away from the status quo in terms of benefits of content.	- This option relies on regulatory sanctions to ensure adherence to regulatory mechanisms applying to potentially harmful content, which may create perverse incentives for MSPs to be overly cautious in moderating or removing content, leading to potential loss of access.
International Alignment Does the option align with what other comparable jurisdictions are doing? Weighting: x1	0 Reflects that of many international jurisdictions, where consumer education and incentivisation initiatives are often executed by a variation of agencies.	+ Better alignment through a focus on education initiatives which reflects some other jurisdictions that spans across all media channels.	+ Better alignment as a proactive and education focused approach is commonly found in comparable jurisdictions.	+ Better alignment as it reflects some of the stronger regulatory measures adopted by other comparable jurisdiction.
Alignment with International Commitments Does the option align with international instruments, treaties, agreements, etc that New Zealand has led, signed on to or expressed support for? Weighting: x1	0 The status quo provides limited opportunities to pursue these.	+ Aligns with international commitments through a coordinated effort to support and educate creators on how content can cause harm.	++ Aligns much better through a range of levers to help educate and support content creators and consumers on the impact of harmful content.	- May not align due to reliance on regulatory sanctions, which may create perverse incentives for MSPs to be overly cautious in moderating or removing content, leading to potential loss of access.
Overall Assessment	0	9	10	1

Section 2.4: Objective 4 - Creators of content (including social media users) are incentivised to minimise the potentially harmful impact of that content

What options are being considered?

166. This objective seeks to motivate creators of content to minimise the potentially harmful impact of their content on consumers or others, given that creators will not be the ones who experience the potentially harmful impacts of their own content.

Option 4A – Status Quo

Analysis

167. The status quo provides a regulatory framework for different kinds of media channels, working indirectly to disincentivise the creation of content that breaches agreed codes and standards. This coverage is fragmented and incomplete, and it does not apply to un-curated user-generated content.
168. Existing codes are relatively adaptable as they can be modified to address emerging issues but remain specific to industry and media channels. The levers under the status quo are generally proportional to risk of harm as most measures are voluntary and industry developed codes that rely on reactive enforcement through complaints from consumers. Most voluntary standards are aimed at enhancing trust in the relevant industry or media channel, which means they tend to increase the trust and safe accessibility of the relevant media.
169. The status quo is largely similar to comparable international jurisdictions however, these generally have more advanced and transparent measures (e.g. more specific voluntary codes with regard to representation of diversity in broadcast media). The status quo does not hinder New Zealand's ability to meet its international commitments but is weak in relation to all online media and very weak in relation to user-generated content.

Option 4B – Supportive approach

Analysis

170. This approach would be largely the same as the status quo in achieving the objective. There would be greater public awareness and minimum standards across all media rather than being constrained by media channel type. However, the levers are much the same as the status quo and as a result would produce the same outcomes.
171. This would be better than the status quo for adaptability because the approach would be comprehensive and not specific to certain media channels. However, there will be some limitations due to partial reliance on Government funding education/awareness initiatives. There would be no change to proportionality to risk of harm as the reliance on education and awareness has few, if any, negative impacts on creators and any overreach would be within the discretion of creators.
172. This would be better than the status quo for maximising benefits of content. This is because it would enable creators, particularly creators of content that is subject to editorial control, to understand and minimise risk of harm from the content they create and maximises opportunities for them to create public interest content and express views and opinions. Education levers may also have a positive impact on the awareness and understanding of user-generated content creators about what could be harmful to different communities and parts of society.
173. International alignment and meeting international commitments would be better than the status quo as there would be a concerted effort to support and educate creators on harm, which aligns with preserving internationally agreed rights to freedom of speech

and access to information. This option also provides more levers to meet its international commitments.

Option 4C – Balanced approach

Analysis

174. Under this approach, creators of content would have greater motivation to be aware and avoid creating content that may cause harm to others, which would be better than the status quo in achieving this objective. This is because there would be government mandated minimum standards for all media channels to increase incentives for creators of content to minimise harmful impact of content, combined with greater public education and awareness of harms amongst user-creators.
175. This would be better than the status quo for adaptability because the approach will be comprehensive and not specific to any media channels. There may be some limitations on adaptability and responsiveness due to partial reliance on Government mandates and funding for education initiatives. However, for proportionality this approach would be the same as the status quo as codes and minimum standards can be developed to provide proportional responses to different degrees of harm.
176. The reliance on industry codes being subject to minimum standards will allow the development of proportionate media specific incentives to minimise risk, while resulting in a safer and more attractive media landscape. This makes the approach better than the status quo for maximising benefits of content.
177. Much like the supportive approach, international alignment better than the status quo as there would be a concerted effort to support and educate creators on harm, which aligns with preserving internationally agreed rights to freedom of speech and access to information. This option would be much better than the status quo in meeting international commitments as it provides a range of different levers for Government and industry to meet its international commitments.

Option 4D – Prescriptive approach

Analysis

178. This approach would be much better than the status quo in achieving the objective as there would be an extension of criminal and civil sanctions to the creation of content that can cause harm. This will significantly increase incentives to avoid creating content that is captured by the criminal and civil sanctions.
179. This approach would be worse than the status quo for adaptability because of the reliance on legislation to create and specify penalties. Although offences that are non-media specific and described in terms of harms impacts will have some inherent flexibility. The reliance on criminal and civil sanctions to disincentivise the creation of content with only moderate risk of harm makes this approach much worse than the status quo. This is because it could create an incentive to be excessively cautious about the creation of content that has a low risk of harm or no harm.
180. This would also be worse than the status quo for maximising benefits of the media. This is due to the risk that regulatory sanctions against content that carries only a moderate risk of harm may disincentivise the creation of some public interest content. This could include opinion pieces and informative podcasts.
181. This approach is also worse than the status quo in alignment with international jurisdictions, as the focus on criminal and civil sanctions as disincentive for creators. This is not aligned with international approaches as there is the potential to lead to loss of freedoms and expression and access to content that has low risks of harm. However, this option is still better than the status quo in meeting international commitments as it provides additional levers to mitigate harms caused by content across all media channels.

How do the options compare to the status quo/counterfactual?

Objective 4: Creators of content (including social media users) are incentivised to minimise the potentially harmful impact of that content

	Option 4A – Status Quo	Option 4B – Supportive approach	Option 4C – Balanced approach	Option 4D – Prescriptive approach
Extent to which objective is achieved (x3) To what extent is the option able to meet the objective? Partially/fully/not at all	0 Existing regulatory frameworks for different kinds of media work indirectly to disincentivise the creation of content that breaches agreed codes and standards. Coverage is fragmented and incomplete and does not apply to un-curated user-generated content. The prohibition on objectionable content disincentivises the creation of that content.	0 This is the same as the status quo in achieving the objective, with a more coordinated focus on creating awareness and minimum standards across all media channels. This is combined with education focused initiatives.	+	++ This achieves the objective through the extension of criminal and civil sanctions to the creation of content with a wider range of harm impacts which will significantly increase incentives to avoid creating that content.
Adaptability (x2) How flexible is the option? For example, how well can this option adapt to different media and content types, future changes in the media landscape, technological advances, changing societal expectations? Evaluate flexibility that's needed to align with cultural values and NZ context	0 Existing codes can be updated and modified to address emerging issues but remain industry/media specific. The statutory definition of objectionable content.	+	+	- Less adaptable because of reliance on legislation to create and specify penalties, although offences that are non-media specific and described in terms of harm impacts will have some inherent flexibility.
Proportionality (x2) Do the expected benefits of the option justify its costs/negative impacts? e.g. would the option result in restriction of access to content that isn't harmful?	0 Because most measures are voluntary industry developed codes that rely on reactive post hoc enforcement through complaints, negative effects are very limited, and trade-offs are managed by MSPs.	0 Proportionality largely the same as reliance on education and awareness has few if any negative impacts and any overreach would be within the discretion of creators.	0	-- Much less proportionate due to reliance on criminal and civil sanctions to disincentivise the creation of content with only moderate risk of harm and relies on judicial discretion to apply sanctions proportionally. May incentivise excessive caution about the creation of low/no risk material.
Maximising benefits of content (x1) To what extent does the option preserve the private and public benefits of content?	0 Most voluntary standards are aimed at enhancing trust in the relevant industry/media, so tend to increase the trust and safe accessibility of relevant media.	++ Enabling creators to understand and minimise the risk of harm from the content they create maximises opportunities for them to create public interest content and express views and opinions.	+	- Some risk that regulatory sanctions against content that carries a moderate risk of harm may disincentivise the creation of some public interest content.
International Alignment (x1) Does the option align with what other comparable jurisdictions are doing?	0 Reflects that of many international jurisdictions, where consumer education and incentivisation initiatives are often executed by a variation of agencies.	+	+	- Worse as regulatory sanctions to incentivise consumers have resulted in over-regulation of potentially harmful content in other jurisdictions, leading to adverse impacts on freedom of expression and a free internet.
Alignment with International Commitments (x1) Does the option align with international instruments, treaties, agreements, etc that New Zealand has led, signed on to or expressed support for?	0 The status quo does not hinder New Zealand to meet its obligations under various instruments and treaties.	+	++	+
Overall Assessment	0	6	9	-1

Section 2.5: Objective 5 – Creators of content (including social media users) are incentivised not to create very harmful content

What options are being considered?

182. This objective seeks incentives for creators to avoid creating very harmful content that are sufficient to outweigh commercial and other incentives to create such content.

Option 5A – Status Quo

Analysis

183. The status quo provides for fragmented government regulatory regimes where different approaches are taken to content on different media, and content on some media is, effectively, not regulated by Government at all. Where content is subject to government regulation, adaptability is limited and constrained to the relevant media channel. Existing government responses are proportionate on the particular media channels that are currently regulated, whereas in the online space the approach to content is inconsistent and may not be proportionate to the risk of harm.
184. The status quo reflects that of many international jurisdictions, where various agencies will provide initiatives for creators, but often these are more advanced and accessible in comparison. The status quo does not hinder New Zealand's ability to meet its obligations under various international treaties and instruments, but neither does it facilitate this.

Option 5B – Supportive approach

Analysis

185. This approach would be worse than the status quo in meeting the objective. Although adaptable, this approach takes a heavy education-focused approach. It moves away from the use of enforcement mechanisms to prevent content creators from producing very harmful content which makes it less likely to achieve this objective than the status quo. The heavier focus on education over enforcement action (such as prosecutions) may not be proportional for very harmful content.
186. As this takes an approach focused on education and supporting industry and community initiatives, there are less regulatory limitations on the creation of content than the status quo.
187. This option is not aligned with stricter measures taken by other comparable jurisdictions that place emphasis on strongly discouraging, and then preventing, very harmful content. This option is consistent with a global movement towards education and the influence this has on the content created, but education alone is not effective at minimising harm in cases where the creator has a deliberate objective of creating very harmful content.
188. It is likely that the levers captured in this option are not 'strong/influential' enough to actively deter the creation of very harmful content, thus posing a risk that not all New Zealand's commitments under international agreements can be effectively met.

Option 5C – Balanced approach

Analysis

189. This approach would be better than the status quo to meet the objective as it enables government prohibition of very harmful content to apply regardless of the media channel the content is created on. The prohibition would be complemented by the use

of mandatory codes to engage MSPs and consumers in the prevention and removal of very harmful content. MSP reporting would inform regular reviews of statutory thresholds for very harmful content, maximising the adaptability of the statutory framework. This approach allows for government regulatory oversight on very harmful content through a range of mechanisms, and this range allows for more proportionate responses.

190. As there will be regulatory oversight of very harmful content on all media channels, there may be impacts on the types of content that creators choose to post to avoid accountability whereas within the status quo they may not have been held accountable and may have otherwise posted the content.
191. This option best reflects the approach that comparable international jurisdictions have taken and does not hinder New Zealand in meeting its obligations under various instruments and treaties.

Option 5D – Prescriptive approach

192. Under this option the creation of very harmful content will be subject to civil or criminal liability. Government would require publishers/providers to take prescribed measures to avoid hosting very harmful content. Non-compliance will be subject to civil or criminal liability.
193. The primary levers under this approach could include:
 - Government criminalising the creation of very harmful content and prosecuting creators of very harmful content. Significant civil and criminal penalties would constitute a strong disincentive to create the content.
 - Government using take-down and removal of content measures to incentivise users to not create very harmful content.
 - Government reserving powers to ensure that industry creators of very harmful content are not able to participate/ are blocked in the New Zealand market.

Analysis

194. This approach would be better than the status quo to meet the objective as it enables government regulation for very harmful content regardless of the media channel the content is created on. However, this may be less adaptable than the status quo as it will involve only top-down Government mandates rather than a mix of legal prohibition and requirements for proactive to industry actions. Reliance solely on stringent regulatory sanctions which preclude more adaptive industry responses to emerging harm.
195. With respect to proportionality, this would be an improvement on the status quo as the creation of very harmful content would capture content creators on all media channels, with a strong focus on government mandates to enforce the prevention of very harmful content. These government mandates may limit freedom of speech/expression, but only where justified and improved trust in the safety of the content environment is likely to improve accessibility and use.
196. This option does not hinder New Zealand's ability to meet its obligations under various instruments and treaties and similar approaches have been taken in France, Germany, the UK and Ireland.

How do the options compare to the status quo/counterfactual?

Objective 5: Creators of content (including social media users) are incentivised not to create very harmful content

	Option 5A – Status Quo	Option 5B – Supportive approach	Option 5C – Balanced approach	Option 5D – Prescriptive approach
Extent to which objective is achieved (x3) To what extent is the option able to meet the objective? Partially/fully/not at all	0 Fragmented Government regulatory regimes mean that different approaches are taken to content on different media, and content on some media is not regulated at all which means that very harmful content on some platforms may not be detected.	- Less able to meet the objective due to heavy education focus and moves away from enforcement mechanisms to incentivise creators to not create very harmful content.	++ Fully achieves the objective as it enables government regulation for very harmful content regardless of the media channel the content is created on using a range of regulatory levers.	++ Partially achieves the objective as it enables government regulation for very harmful content regardless of the media channel the content is created on through Government mandates and enforcement measures.
Adaptability (x2) How flexible is the option? For example, how well can this option adapt to different media and content types, future changes in the media landscape, technological advances, changing societal expectations? Evaluate flexibility that's needed to align with cultural values and NZ context	0 In the areas where content is subject to Government regulation, adaptability is limited and constrained by media channel. There are voluntary approaches which undergo regular review to reflect and adapt to the contemporary operating environment, which may capture creators of very harmful online content.	++ Highly adaptable due to focus being on industry and community-led initiatives (with Government support) that enable flexibility of approach, particularly to changing nature of harms and societal expectations.	+ Adaptable as Government, as well as criminalising very harmful content, mandates industry to take a proactive approach. Allows some flexibility in approach to respond to changing nature of very harmful content.	0 Equally adaptable This would have similar adaptability to the status quo as it relies solely on government regulation and enforcement.
Proportionality (x2) Do the expected benefits of the option justify its costs/negative impacts? e.g. would the option result in restriction of access to content that isn't harmful?	0 Existing Government responses are proportionate to the particular media channels that are currently regulated. Online content is not regulated by Government, which means the ways that harmful content is addressed by different MSPs are inconsistent and may not be proportionate to the risk of harm.	- Less proportional due to focus on education over enforcement which may not be appropriate for very harmful content.	++ High proportionality as it would capture content creators on all media channels and allows for government regulatory oversight on very harmful content through a range of mechanisms against very harmful content.	+ Proportional as the creation of very harmful content would capture content creators on all media channels, with a strong focus on government mandates to enforce the prevention of very harmful content being created.
Maximising benefits of content (x1) To what extent does the option preserve the private and public benefits of content?	0 Status quo maximises freedom of speech and creation of content on platforms that are not currently regulated by Government, outside of <u>objectionable</u> content.	+ As this takes an approach focused on education and supporting industry and community initiatives, there is less regulatory limitations on the creation of content than the status quo.	0 Some reduction of access to benefits of content from regulation of content on all media channels. However, this would offset by safer access for people who would have otherwise experienced harms from very harmful content.	0 Some reduction of access to benefits of content from regulation of content on all media channels. However, this would offset by safer access for people who would have otherwise experienced harms from very harmful content.
International Alignment (x1) Does the option align with what other comparable jurisdictions are doing?	0 Reflects that of many international jurisdictions, where consumer education and incentivisation initiatives are often executed by a variation of agencies.	+ Better alignment through a focus on education initiatives which reflects some other jurisdictions that spans across all media channels.	++ Better alignment as a mixture of strong regulation and proactive industry approach is commonly found in comparable jurisdictions.	- Worse as regulatory sanctions to incentivise consumers have resulted in over-regulation of potentially harmful content in other jurisdictions, leading to adverse impacts on freedom of expression and a free internet.
Alignment with International Commitments (x1) Does the option align with international instruments, treaties, agreements, etc that New Zealand has led, signed on to or expressed support for?	0 The status quo does not hinder New Zealand to meet its obligations under various instruments and treaties.	+ Aligns with international commitments through a coordinated effort to support and educate creators on how content can cause harm.	++ Aligns much better than the status quo through a range of levers to prevent very harmful content coupled with initiatives to mandate industry to take a proactive approach to preventing harmful content.	++ Aligns better than the status quo through an extension of civil and criminal sanctions to creators of very harmful content.
Overall Assessment	0	2	16	9

Section 2.6: Objective 6 – Publishers/providers of media and online content are required to minimise the prevalence and impact of potentially harmful content

What options are being considered?

197. This objective seeks clear authority and obligations for MSPs to take action to reduce potential harm from content.

Option 6A – Status Quo

Analysis

198. For content that is covered by the existing legislative framework, there are co-regulatory or industry-led (voluntary) codes and standards for minimising the prevalence and impact of potentially harmful content. This includes age-based consumer warnings for certain media channels, such as television and cinematic releases. Providers of online content are largely outside of Government regulation and are not mandated to minimise the prevalence and impact of potentially harmful content.
199. The status quo has limited adaptability as it is constrained by which media channel the content is found on, however voluntary approaches undergo regular review to reflect and adapt to the contemporary operating environment. The status quo is also proportionate to the particular media channels that are currently regulated. However, content not captured such as online content is left for publishers/providers to respond to which may not be proportionate to the risk of harm and can be inconsistent. The status quo does maximise freedom of speech and access on services of publishers/providers that are not currently regulated by Government, although this may inhibit freedom of people who experience harm arising from the unregulated nature of these platforms.
200. In terms of international alignment, the status quo has several disparities to comparable jurisdictions, where content is largely enforced by a single expert agency. Fragmented Government regimes also mean that there are inconsistent approaches to different content. There is also a risk that harmful content on some platforms may not be detected and the platforms not being held to account which needs to be improved on to meet international commitments.

Option 6B – Supportive approach

Analysis

201. This approach is better than the status quo in achieving the objective as content would be covered across all media channels. This is done through Government developing minimum standards for minimising harm and facilitates/supports publishers/providers to adopt these minimum standards. This may include education or public awareness campaigns that publishers/providers are responsible for the dissemination of to users of their services, to minimise the prevalence and impact of potentially harmful content.
202. This would be much better than the status quo in adaptability because it relies on industry and community-led initiatives, with Government playing a supporting role. This would enable flexibility in how publishers/providers take action to reduce potentially harmful content on their platforms as the need arises. However, this option would be much worse than the status quo in proportionality to the risk of harm because it does not require publishers/providers to adopt minimum requirements. The focus is on encouraging and educating. With the absence of Government mandates on publishers/providers, some will not operate with minimum standards in place, while others may implement standards inconsistently or ineffectively.
203. This approach would be worse than the status quo for maximising the benefits of content. This is because it would remove existing Government requirements on

publishers/providers to reduce potential harm from content that are captured in the status quo by legislation such as cinematic films. While this would enhance freedom of expression and access to content, this would mean that consumers are at risk of harm on more media services. This includes those that would have otherwise been operated with codes and standards that focused on, or contributed to, harm minimisation.

204. This would be better than the status quo for international alignment by reflecting the education focus taken by some comparable jurisdictions. However, the Government would only play a supporting role by encouraging industry and moves away from setting requirements on publishers/providers and using enforcement mechanisms. This means there is more likely to be adverse impacts on the rights of marginalised communities to be free from harm and discrimination, making this approach worse than the status quo for meeting international commitments.

Option 6C – Balanced approach

Analysis

205. This approach is much better than the status quo at achieving the objective through the Government setting minimum standards for harm minimisation from content across all media channels. This can include minimum performance standards for publishers/providers. Government would facilitate and support publisher/provider-led development of harm minimisation codes for their respective media services. Government would also encourage industry-led initiatives to educate consumers about potentially harmful content. This provides a range of levers to allow for Government to mandate publishers/providers to minimise the prevalence and impact of potentially harmful content.
206. Adaptability would be better than the status quo as Government enables publishers/providers or their industry representative bodies to develop their own harm minimisation codes to comply with required minimum standards. This allows the flexibility for publishers/providers to implement minimum requirements in ways that are compatible with their operating systems and what they may already be doing in other international jurisdictions. The result of publishers/providers being able to develop their own harm minimisation codes makes this approach much better than the status quo for proportionality. This is because the flexibility allows for harm minimisation requirements to match different types of content and the potential harms that can come from the content. Government has the flexibility to respond to regulatory non-compliance through different regulatory and enforcement tools and is able to scale responses depending on the severity and risk of harm from such non-compliance.
207. This approach would be better than the status quo in maximising the benefits of the media because of the flexibility that comes with Government enabling publishers/providers to develop their own codes but still providing required minimum standards supported by regulatory enforcement tools for compliance. The flexibility means maximising benefits of media content for a larger proportion of consumers, including those who may otherwise be subject to discriminatory content.
208. This option is much better than the status quo for international alignment. Under this option, Government sets minimum standards for harm minimisation while enabling publishers/providers to develop their own codes to comply with the minimum requirements. This option also sees government facilitating industry education initiatives and making use of consumer warnings. This is in line with industry-led code-based approaches that are being adopted or developed by comparable jurisdictions, and regulatory enforcement measures that these jurisdictions have adopted.

Option 6D – Prescriptive approach

Analysis

209. This approach is much better than the status quo in achieving the objective. This is because the Government would develop harm minimisation codes and mandate publishers/providers to adopt and implement these codes, which would cover across all media channels. Publishers/providers will be subject to criminal or civil penalties for systemic non-compliance with the mandated harm minimisation codes.
210. This would be worse than the status quo for adaptability as it relies on Government mandates for publishers/providers to comply with codes. This makes the approach less adaptable as it would require going through the parliamentary processes which can be time consuming. This means the approach cannot quickly address emerging harms, changing societal expectations and better alignment with New Zealand's cultural values.
211. In terms of proportionality, it would also be worse than the status quo as it relies on criminal or civil penalties for systemic non-compliance, which may create perverse incentives for publishers/providers to be overly cautious in moderation or removing content on their platforms that poses low risk or no risk of harm. Due to the potential for creating this perverse incentive, this option would be much worse than the status quo in maximising the benefit of content by negatively impacting on freedom of access and provision of information.
212. However, this approach is better than the status quo for international alignment and meeting international commitments. The approach includes levers taken in comparable jurisdictions such as Germany and France, where the removal of potentially harmful content is a top priority which can potentially compromise aspects of freedom of expression.

How do the options compare to the status quo/counterfactual?

Objective 6: Publishers/providers of media and online content are mandated to minimise the prevalence and impact of potentially harmful content

	Option 6A – Status Quo	Option 6B – Supportive approach	Option 6C – Balanced approach	Option 6D – Prescriptive approach
Extent to which objective is achieved (x3) To what extent is the option able to meet the objective? Partially/fully/not at all	0 Content covered by the legislative framework, there are co-regulatory or industry-led (voluntary) codes and standards for minimising the prevalence and impact of potentially harmful content. Providers of online content are largely outside of Government regulation and are not mandated to minimise the prevalence and impact of potentially harmful content.	+	++	++
Adaptability (x2) How flexible is the option? For example, how well can this option adapt to different media and content types, future changes in the media landscape, technological advances, changing societal expectations? Evaluate flexibility that's needed to align with cultural values and NZ context	0 In the areas where content is subject to Government regulation, adaptability is limited and constrained by media channel. There are voluntary approaches which undergo regular review to reflect and adapt to contemporary operating environment.	++	+	0
Proportionality (x2) Do the expected benefits of the option justify its costs/negative impacts? e.g. would the option result in restriction of access to content that isn't harmful?	0 Existing Government responses are proportionate to the particular media channels that are currently regulated. Online content is not regulated by Government, which means inconsistencies to how harmful content is addressed.	--	++	-
Maximising benefits of content (x1) To what extent does the option preserve the private and public benefits of content?	0 Status quo maximises freedom of speech and access on the services of publishers/providers that are not currently regulated by Government. However, this may inhibit freedoms of people or groups who experience harm arising from the unregulated nature of these platforms.	-	+	--
International Alignment (x1) Does the option align with what other comparable jurisdictions are doing?	0 Several disparities to regulation compared to other jurisdictions but emphasis on freedom of expression and acknowledgement that users are experiencing harm is consistent.	+	++	+
Alignment with International Commitments (x1) Does the option align with international instruments, treaties, agreements, etc that New Zealand has led, signed on to or expressed support for?	0 The status quo supports New Zealand meeting its obligations under various instruments and treaties in respect of some very harmful material, but existing regulation has incomplete coverage.	-	+	+
Overall Assessment	0	2	16	4

Section 2.7: Objective 7 – Publishers/providers of media and online content are incentivised not to promote potentially harmful content

What options are being considered?

213. This objective seeks incentives for publishers/providers to avoid promoting potentially harmful content (e.g. by algorithmic suggestions, recommendations, search result prioritisation) that are sufficient to outweigh commercial and other incentives to promote such content.

Option 7A – Status Quo

Analysis

214. The status quo does not include specific Government initiatives to address the promotion of potentially harmful content. There is prohibition on possession, distribution and advertising of objectionable material which may have some disincentive effect in relation to the promotion of very harmful content. There is limited adaptability as this disincentive is dependent on the current statutory definition of ‘objectionable’ (this covers very harmful content). Proportionality to risk of harm is not an issue as the impact of this lever applies only to very (including extremely) harmful content.

Option 7B – Supportive approach

Analysis

215. This approach would be slightly better than the status quo in achieving the objective, through encouraging specific measures to discourage the promotion of potentially harmful content through codes and education. But against this, there is strong commercial incentives that will continue to exist to maximise audience engagement through the promotion of high impact content which are likely to be potentially harmful.

216. This approach would be much better than the status quo for adaptability due to the reliance on education measures and Government playing a supporting/encouraging role. This allows industry initiatives to be highly adaptable as they would not rely on Government mandates. This adaptability would also allow for appropriate proportionality to lower risks of harm, because reliance on education measures means initiatives would likely have moderate costs and limited negative effects.

217. As no content is made unavailable by the levers under this approach, there would be more nuanced approaches to the promotion of content. This means it would be similar to the status quo in terms of impact on the benefits of the media. This would also be similar to the status quo in aligning with international jurisdictions. It would be slightly better than the status quo in aligning with international commitments to address harmful impacts arising from algorithmic content programming.

Option 7C – Balanced approach

Analysis

218. This approach would be much better than the status quo in achieving the objective as it would require specific measures to reduce the promotion of potentially harmful content, which could build on and align with existing initiatives by platforms and industries. This would be done through codes. The creation of a specific offence for promoting very harmful content would reinforce the importance of proactive monitoring of what content is promoted on platforms.

219. The requirement to address this issue in mandatory codes would be much more adaptable than the status quo through reliance on industry to determine how codes could discourage the promotion of potentially harmful content. This means the

approach would also be better for proportionality, as differing approaches could be tailored to avoid promoting harmful content with minimum disruption to the operating model of each MSP.

220. This would be similar to the status quo in terms of the benefits of content. No content would be made unavailable by these measures. Loss of the access benefits of algorithmic suggestions and prioritisation would be limited to potentially harmful content and offset by increased safety for consumers.
221. In terms of international alignment and meeting international commitments, this approach would be much better than the status quo. Other jurisdictions are also taking code-based approaches to deal with algorithmically promotion of harmful content. Targeting algorithmic promotion of violent extremist and racist content is a key element of Christchurch Call commitments.

Option 7D – Prescriptive approach

Analysis

222. This approach would be much better than the status quo in achieving the objective by the creation of a specific offence for promotion of potentially harmful content (over a specified threshold). This would constitute a big disincentive for publishers/providers of content to avoid promoting potentially harmful content as they could be subject to civil and criminal liability for doing so. Further disincentives include the possibility of exclusion from the market for repeated non-compliance.
223. This could be better than the status quo for adaptability if the threshold is prescribed in a flexible way. While the approach has a reliance on statutory offences, these offences can be worded to be type of harm neutral in order to be more adaptable. This would also be better than the status quo for proportionality as there would be measures in place for a range of potentially harmful content, but no content is made unavailable by these measures.
224. As no content is made unavailable through this approach, there is no negative impact on freedom of expression and can therefore be better than the status quo. However, this would impact on ease of access to some types of content, and there may be significant compliance costs for MSPs.
225. This approach would be better than the status quo in terms of both international alignment and meeting international obligations. The approach would align with comparable jurisdictions such as Germany and France where the removal of potentially harmful content is a priority and preventing the promotion of potentially harmful content is important. This option would also allow for stronger enforcement actions through Government regulation, which could enable New Zealand to better meet international commitments Targeting algorithmic promotion of violent extremist and racist content is a key element of Christchurch Call commitments.

How do the options compare to the status quo/counterfactual?

Objective 7: Publishers/providers of media and online content are incentivised not to promote potentially harmful content

	Option 7A – Status Quo	Option 7B – Supportive approach	Option 7C – Balanced approach	Option 7D – Prescriptive approach
Extent to which objective is achieved (x3) To what extent is the option able to meet the objective? Partially/fully/not at all	0 No current Government initiatives specifically address the promotion of potentially harmful content. The prohibition on possession, distribution and advertising of objectionable material may have some disincentive effect.	++ No better or worse than the status quo in achieving the objective as strong commercial incentives will continue to exist to maximise audience engagement through the promotion of high impact content.	++ Achieves the objective by requiring specific measures to discourage the promotion of potentially harmful content through codes and the creation of a specific offence for promotion of very harmful content.	++ Achieves the objective by the creation of a specific offence for promotion of potentially harmful content as a disincentive.
Adaptability (x2) How flexible is the option? For example, how well can this option adapt to different media and content types, future changes in the media landscape, technological advances, changing societal expectations? Evaluate flexibility that's needed to align with cultural values and NZ context	0 The limited current disincentive is based on the statutory definition of objectionable.	++ Highly adaptable due to reliance on education measures and Government encouragement of industry initiatives.	++ High adaptability due to reliance on industry developed codes to discourage the promotion of potentially harmful content provides a high degree of adaptability. Combined with reliance on statutory offences to address the promotion of very harmful content may be less so.	++ Adaptable due to reliance on statutory offences to address the promotion of all potentially harmful content is less agile, but the offence can be worded to be type of harm neutral.
Proportionality (x2) Do the expected benefits of the option justify its costs/negative impacts? e.g. would the option result in restriction of access to content that isn't harmful?	0 The limited current disincentive is based on the statutory definition of objectionable which applies only to very harmful content	++ Highly proportional due to reliance on education measures and Government encouragement of industry initiatives would likely have moderate costs and limited negative effects.	++ Highly proportional as discouraging the promotion of all potentially harmful content, and criminalising the promotion of very harmful content, may disrupt the business models of some MSPs work as an incentive. Because no content is made unavailable by these measures, there is no negative impact on freedom of expression.	++ Limited proportionality as prohibiting the promotion of all potentially harmful content will significantly disrupt the business models of MSPs. No content is made unavailable by these measures so there is no negative impact on freedom of expression, although there may be significant impact on ease of access.
Maximising benefits of content (x1) To what extent does the option preserve the private and public benefits of content?	0 The limited current disincentive is based on the statutory definition of objectionable which applies only to publications "likely to be injurious to the public good"	0 No content is made unavailable by these measures and the absence of Government mandates may allow more nuanced approaches to the promotion of content.	++ No content is made unavailable by these measures so there is no negative impact on freedom of expression, although there may be some impact on ease of access to some content. Against this, trust in the safety of the internet and such recommendations as remain will be enhanced.	++ No content is made unavailable by these measures so there is no negative impact on freedom of expression, although there may be some impact on ease of access to some content. Against this, trust in the safety of the internet and such recommendations as remain will be enhanced.
International Alignment (x1) Does the option align with what other comparable jurisdictions are doing?	0 Several disparities to regulation compared to other jurisdictions but emphasis on freedom of expression and acknowledgement that users are experiencing harm is consistent.	0 Little alignment as comparable jurisdictions are also utilising a range of education-based levers.	++ Much better alignment as comparable jurisdictions utilise a range of different levers for different levels/types of harms from content, seeking to help consumers avoid harmful content.	++ Better alignment as comparable jurisdictions have been adopting stronger Government intervention mechanisms to respond to a diversifying range of potentially harmful content.
Alignment with International Commitments (x1) Does the option align with international instruments, treaties, agreements, etc that New Zealand has led, signed on to or expressed support for?	0 The prohibition on possession, distribution and advertising of objectionable material is aligned to obligations under the Christchurch Call and other international obligations to combat discrimination.	++ Better aligns with international commitments as government plays a supporting role by encouraging industry to avoid promoting harmful content on their platform. There would be some scope to give effect to international obligation in this way.	++ Much better alignment as it enables government regulation for very harmful content regardless of the media channel and would require industry initiatives to avoid promoting harmful content at all levels of harm. This would enable stronger action to meet international obligations.	++ Better alignment as it enables government regulation for potentially harmful content regardless of the media channel the content is created on. This is done through high sanctions for non-compliance. This would enable stronger action to meet international obligations.
Overall Assessment	0	12	19	13

Section 2.8: Objective 8 – Publishers/providers of media and online content are incentivised not to publish/provide very harmful content

What options are being considered?

226. This objective seeks incentives to avoid publishing/providing very harmful content that are sufficient to outweigh commercial and other incentives to publish/provide such content.

Option 8A – Status Quo

Analysis

227. The status quo provides for fragmented government regulatory regimes where different approaches are taken to content on different media, and content on some media is, effectively, not regulated at all while platforms are held to no account despite hosting content that is very harmful.

228. In the areas where content is subject to government regulation, adaptability is limited and constrained by media channel. There are voluntary approaches which undergo regular review to reflect and adapt to contemporary operating environment; these may encourage publishers/platforms to self-regulate to ensure they do not host very harmful content on their platforms. Existing government responses are proportionate to the particular media channels that are currently regulated, whereas in the online space the approach to content is inconsistent and may not be proportionate to the risk of harm.

229. There is alignment in the types of content that are considered very harmful, and an understanding that, given online content's borderless nature, an internationally cohesive approach may be required to build a safer internet. Overseas there are increasing moves for a single Regulator to oversee the classification and removal of very harmful content, whereas in New Zealand it is spread across multiple agencies. The status quo does not hinder New Zealand's ability to meet its obligations under various international treaties and instruments.

Option 8B – Supportive approach

Analysis

230. Under this approach government plays a supporting role by encouraging industry to reduce demand and publication of very harmful content on their platform. It moves away from the use of enforcement mechanisms requiring platforms to take-down very harmful content which makes it less likely to achieve this objective than the status quo.

231. There is a high degree of flexibility as there is a focus on working with industry by encouraging MSPs to reduce demand and publication of very harmful content on their platforms. This option takes away the existing enforcement mechanisms as it will be largely industry led in the reduction of demand and publication of very harmful content on their platforms and so would not be proportional to very harmful content. There are less regulatory limitations on the publication of content than with the status quo.

232. This approach is not aligned with international jurisdictions, where an enforcement heavy approach is taken to content that is considered to be very harmful. However, levers included under these options are also popular as a mitigative step in reducing the scale and engagement with such content. This option involves moving away from the use of enforcement mechanisms for very harmful content which would constrain New Zealand from meeting some international commitments, such as the Christchurch Call.

Option 8C – Balanced approach

Analysis

233. This option is an improvement on the status quo in achieving the objective as it enables government regulation for very harmful content regardless of the media channel. It does this by working with industry to take proactive approaches to preventing the upload/hosting of very harmful content. Government mandates MSPs/platforms to adopt and implement codes and practices to reduce demand and opportunities for very harmful content. This allows for adaptability to the changing nature of content and what is considered very harmful to be adopted in the implementation of industry protocols and practices.
234. This approach limits the amount of very harmful content that people are able to publish. However, this may have benefits for people who would have otherwise experienced harms from very harmful content.
235. This approach is much more heavily aligned with other jurisdictions in which there are greater enforcement powers and there is a movement towards more centralised regulation and enforcement of this very harmful content. This option does not hinder New Zealand in meeting its obligations under various instruments and treaties.

Option 8D – Prescriptive approach

Analysis

236. This option is an improvement on the status quo in achieving the objective as it enables government regulation for very harmful content regardless of the media channel the content is created on, through high sanctions for non-compliance. This may be less adaptable than the status quo as it will enable government to prescribe types of content that must not be published with very high criminal or civil penalties for non-compliance and such sanctions may be time consuming to change.
237. With respect to very harmful content, this approach would be an improvement on the status quo as the publication of very harmful content would be captured on all media channels. This would impact on freedom of expression as industry may be more reluctant to publish content that may not be very harmful at the risk of facing very high criminal or civil penalties for non-compliance, although this may have benefits for people who would have otherwise experienced harms from very harmful content.
238. This approach is more heavily aligned with other jurisdictions in that there are greater enforcement powers and is also aligned with comparable jurisdictions with respect to mandated transparency reporting. This option does not hinder New Zealand's ability to meet its obligations under various instruments and treaties.

How do the options compare to the status quo/counterfactual?

Objective 8: Publishers/providers of media and online content are incentivised not to publish/provide very harmful content

	Option 8A – Status Quo	Option 8B – Supportive approach	Option 8C – Balanced approach	Option 8D – Prescriptive approach
Extent to which objective is achieved (x3) To what extent is the option able to meet the objective? Partially/fully/not at all	0 Fragmented Government regulatory regimes mean that different approaches are taken to content on different media, and content on some media is not regulated at all which means that very harmful content on some platforms may not be detected and the platforms not being held to account.	- Less able to achieve objective as Government plays a supporting role and moves away from the use of enforcement mechanisms requiring platforms to take-down very harmful content.	++ Achieves the objective as it enables government regulation for very harmful content regardless of the media channel. It does this by working with industry to take proactive approaches to preventing the upload/hosting of very harmful content.	+ Partially achieves the objective as it enables government regulation for very harmful content regardless of the media channel the content is created on. This is done through high sanctions for non-compliance.
Adaptability (x2) How flexible is the option? For example, how well can this option adapt to different media and content types, future changes in the media landscape, technological advances, changing societal expectations? Evaluate flexibility that's needed to align with cultural values and NZ context	0 In the areas where content is subject to Government regulation, adaptability is limited and constrained by media channel. There are voluntary approaches which undergo regular review to reflect and adapt to contemporary operating environment.	++ Highly adaptable as focus is on working with industry by encouraging MSPs to reduce demand for and publication of very harmful content on their platforms. This would enable some flexibility in prioritisation in response to the changing nature and incidence of harms and societal expectations.	+ Adaptable as Government works with industry to enable a proactive approach to preventing the upload of very harmful content. This allows for adaptability to the changing nature of content and what is considered very harmful to be adopted in the implementation of industry protocols and practices.	- Less adaptable as it will enable government to prescribe types of content that must not be published with very high criminal/civil penalties for non-compliance. These are stringent regulatory sanctions which may be time consuming to change.
Proportionality (x2) Do the expected benefits of the option justify its costs/negative impacts? e.g. would the option result in restriction of access to content that isn't harmful?	0 Existing Government responses are proportionate to the particular media channels that are currently regulated. Online content is not regulated by Government, which means publishers/providers of very harmful online content are not held to account.	-- Not very proportional as Government takes a supporting role and may reduce existing enforcement mechanisms as it will be largely industry led. This would not be proportional to very harmful content.	++ High proportionality as it allows for government to mandate MSPs very harmful content on their platforms through a range of monitoring and enforcement mechanisms to ensure MSPs that publish/host very harmful content are held to account if they are non-compliant.	+ Proportional as the publication of very harmful content would be captured on all media channels. MSPs would be held to account through government imposing very high criminal/civil penalties for non-compliance.
Maximising benefits of content (x1) To what extent does the option preserve the private and public benefits of content?	0 Status quo maximises freedom of speech and creation of content on platforms that are not currently regulated by Government, outside of <u>objectionable</u> content.	+ As this takes an approach focused on supporting and encouraging industry to reduce demand and publication of very harmful content on their platforms. There are less regulatory limitations on the publication of content than the status quo.	- As there will be government mandates on MSPs/Platforms for very harmful content on all media channels. This limits the amount of very harmful content that people are able to publish, impacting on freedom of expression.	- This may impact on ability to share and access content if industry becomes more reluctant to publish content that may not be very harmful at the risk of facing very high criminal/civil penalties for non-compliances.
International Alignment (x1) Does the option align with what other comparable jurisdictions are doing?	0 There is alignment in the types of content that are considered very harmful, some of which are criminalised.	- Less alignment due to move away from enforcement approach which is common across comparable jurisdictions in response to very harmful content.	+ More aligned in that there are greater enforcement powers, and there is a movement towards more centralised regulation and enforcement of this content.	++ More heavily aligned in that there are greater enforcement powers. Also aligned with comparable jurisdictions particularly when it comes to mandated transparency reporting.
Alignment with International Commitments (x1) Does the option align with international instruments, treaties, agreements, etc that New Zealand has led, signed on to or expressed support for?	0 The prohibition on possession, distribution and advertising of objectionable material is aligned to obligations under the Christchurch call and other international obligations to combat discrimination.	- Less aligned as Government plays a supporting role and moves away from the use of enforcement mechanisms for very harmful content which constrains NZ meeting some international commitments such as the Christchurch Call.	++ Much better alignment as it enables government regulation for very harmful content regardless of the media channel. This would enable stronger action to meet international obligations.	+ Better alignment as it enables government regulation for very harmful content regardless of the media channel the content is created on. This is done through high sanctions for non-compliance. This would enable stronger action to meet international obligations.
Overall Assessment	0	-4	14	5

Section 2.9: Objective 9 – Consumers are prevented from experiencing extremely harmful content

What options are being considered?

This set of options have been developed for objective

239. This objective seeks to avoid the immediate and severe harm that can be caused by experiencing the most extreme content. This includes content such as CSAM and violent extremist or terrorist material.

Option 9A – Status Quo

Analysis

240. The status quo provides for fragmented government regulatory regimes where different approaches are taken to content on different media, and content on some media is, effectively, not regulated at all. In the areas where content is subject to government regulation, adaptability is limited and constrained by media channel. The definition of 'objectionable' in the Classification Act, which covers most types of extremely harmful content, cannot be easily amended to reflect changing societal expectations and a changing array of harms.
241. Existing mechanisms for extremely harmful content are proportionate to and reflect the extremely high risk of harm. Preventing access of extremely harmful content enables safer use of media.
242. Although the status quo is aligned, internationally, with the types of content that are considered extremely harmful, it is not aligned with other jurisdictions in that these powers are not aligned and coordinated with initiatives to deal with less extreme content. The status quo does not hinder New Zealand to meet its obligations under various instruments and treaties.

Option 9B – Supportive approach

Analysis

243. This approach would be worse than the status quo in meeting the objective because the focus on supporting industry-led initiatives and education is disproportionately negative. It is a flexible approach as it focuses on supporting industry-led initiatives so approaches could change without requiring legislative changes. However, it would seriously hinder the ability of Government to remove content in a timely manner, and the absence of sanctions would mean there are no strong disincentives to the creation or provision of extreme content.
244. This option does not hinder New Zealand's ability to maintain its international commitments, as extremely harmful content is heavily enforced globally, and the benefits of freedom of expression (which some may argue are compromised when content is banned) are outweighed by the imminent harm such content is likely to cause on individual's and societies. The definition of extremely harmful content would need to be reviewed regularly to ensure it reflects national and international risk thresholds.

Option 9C – Balanced approach

Analysis

245. This approach would be better than the status quo to meet the objective as there would be a broader range of response levers for a Regulator to draw upon both in preventing consumer access to extremely harmful content, and a faster response process.

Additionally, this option opens up future opportunities to increase the adaptability of the system due to the broad range of levers available.

246. This option would not be better or worse than the status quo in terms of proportionality and is more heavily aligned with other jurisdictions as there are greater enforcement powers.
247. Like option 9B, this option does not hinder New Zealand's ability to maintain its international commitments, as extremely harmful content is heavily enforced globally, and the benefits of freedom of expression (which some may argue are compromised when content is banned) are outweighed by the imminent harm such content is likely to cause on individuals and societies. The definition of extremely harmful content would need to be reviewed regularly to ensure it reflects national and international risk thresholds.

Option 9D – Prescriptive approach

Analysis

248. This approach would be better than the status quo in meeting the objective, as Government prescribes and monitors MSPs' adoption and implementation of protocols and practices to reduce demand and opportunities for accessing extremely harmful content. This approach also requires other mitigation approaches to prevent MSPs and platforms acting as a vector for extremely harmful content being accessed, shared or widely disseminated. Criminal and civil sanctions will apply to any breach. It uses mandated transparency reporting of industry compliance to monitor the effectiveness of government initiatives. Mandatory transparency reporting would mean that Government is continuously up to date on the effectiveness of approaches to extremely harmful content.
249. This option could result in online platforms being blocked for repeated breaches but this may not be proportionate where the majority of content hosted on a platform does not fall into the extremely harmful content category. As a result of this, this approach could likely impact on people's access to a free and open internet.
250. This approach is more heavily aligned with the approaches of other jurisdictions as there are greater enforcement powers than under the status quo. Similarly, it is more like comparable jurisdictions with respect to mandated transparency reporting. Like options 9B and 9C, this option does not hinder New Zealand's ability to maintain its international commitments, as extremely harmful content is heavily enforced globally, and the benefits of freedom of expression (which some may argue are compromised when content is banned) are outweighed by the imminent harm such content is likely to cause on individual's and societies. The definition of extremely harmful content would need to be reviewed regularly to ensure it reflects national and international risk thresholds.

How do the options compare to the status quo/counterfactual?

Objective 9: Consumers are prevented from experiencing extremely harmful content

	Option 9A – Status Quo	Option 9B – Supportive approach	Option 9C – Balanced approach	Option 9D – Prescriptive approach
Extent to which objective is achieved (x3) To what extent is the option able to meet the objective? Partially/fully/not at all	0 The creation, possession, publication, and distribution of extremely harmful content is subject to civil or criminal liability. Recent amendments to the Classification Act with regard to interim assessments mean that extremely harmful content can be removed immediately without the need for a prior detailed assessment.	-- Does not achieve the objective as Government focuses on supporting industry initiatives to not host and to remove extremely harmful content from their platforms. Government supports community to educate members on the dangers of experiencing extremely harmful content.	++ Achieves the objective as Government mandates MSPs to adopt and implement protocols and practices which will reduce demand and opportunities for accessing extremely harmful content. Government uses a range of levers against extremely harmful content being accessed and disseminated.	+ Partially achieves the objective as Government prescribes and monitors MSP adopting and implementing protocols and practices. Criminal and civil sanctions will apply to any breach. It also uses mandated transparency reporting of industry compliance to monitor the effectiveness of government initiatives.
Adaptability (x2) How flexible is the option? For example, how well can this option adapt to different media and content types, future changes in the media landscape, technological advances, changing societal expectations? Evaluate flexibility that's needed to align with cultural values and NZ context	0 Extremely harmful content is subject to Government regulation, so adaptability is limited and constrained by media channel. The definition of 'objectionable' in the Classification Act, which covers most types of extremely harmful content, cannot be easily amended to reflect changing societal expectations and a changing array of harms.	++ Highly adaptable as this option focuses on supporting industry-led initiatives; therefore, approaches could change without requiring legislative changes. Education is built around what is important to New Zealanders at that point in time, so would be more adaptable to changing cultural values.	0 This option opens up a broader range of levers that still draw upon existing concepts and approaches.	+ Adaptable as there would be mandatory transparency reporting which mean that Government is continuously up to date on the effectiveness of approaches to extremely harmful content.
Proportionality (x2) Do the expected benefits of the option justify its costs/negative impacts? e.g. would the option result in restriction of access to content that isn't harmful?	0 Existing mechanisms for extremely harmful content are proportionate to and reflect the extremely high risk of harm.	- This option is worse than the status quo because the focus on supporting industry-led initiatives and education is disproportionately negative.	0 This option would not be better or worse than the status quo in terms of proportionality.	- This option could result in online platforms being blocked for repeated breaches. This may not be proportionate where the majority of content hosted on a platform does not fall into extremely harmful content.
Maximising benefits of content (x1) To what extent does the option preserve the private and public benefits of content?	0 Existing mechanisms for extremely harmful content are proportionate to and reflect the extremely high risk of harm. Preventing access of extremely harmful content enables safer use of media.	0 This option would not be better or worse than the status quo in terms of maximising benefits of the media.	0 This option would not be better or worse than the status quo in terms of maximising benefits of the media.	-- Significantly limits benefits of content as it could result in online platforms being blocked for repeated breaches. This will impact on people's access to a free and open internet.
International Alignment (x1) Does the option align with what other comparable jurisdictions are doing?	0 There is alignment in the types of content that are considered extremely harmful, and joint processes to prevent access to extremely harmful content (e.g. CSAM).	+ More aligned in that there are more coordinated efforts to support industry-led initiatives and education about extremely harmful content.	+ More aligned in that there are greater enforcement powers.	++ More heavily aligned in that there are greater enforcement powers. Also aligned with comparable jurisdictions particularly when it comes to mandated transparency reporting.
Alignment with International Commitments (x1) Does the option align with international instruments, treaties, agreements, etc that New Zealand has led, signed on to or expressed support for?	0 The status quo does not hinder New Zealand to meet its obligations under various instruments and treaties. Under recent changes to the Classification Act to bring livestreamed content within the Act, extremely harmful content that is livestreamed can be made 'objectionable' (illegal), which aligns with Christchurch Call commitments.	+ This option does not hinder NZs ability to maintain its international commitments, as extremely harmful content is heavily enforced globally, and the benefits of freedom of expression are outweighed by the imminent harm such content is likely to cause on individual's and societies.	+ This option does not hinder NZs ability to maintain its international commitments, as extremely harmful content is heavily enforced globally, and the benefits of freedom of expression are outweighed by the imminent harm such content is likely to cause on individual's and societies.	+ This option does not hinder NZs ability to maintain its international commitments, as extremely harmful content is heavily enforced globally, and the benefits of freedom of expression are outweighed by the imminent harm such content is likely to cause on individual's and societies.
Overall Assessment	0	-2	8	4

Section 3: Preferred option to address the policy problem

What option is likely to best address the problem, meet the policy objectives, and deliver the highest net benefits?

251. As covered in section two, the scope of the options for each objective are set at a high level and aligned with three different co-regulatory approaches. These approaches places emphasis on different regulatory and non-regulatory levers. Based on the analysis of the options identified for each of the objectives in sections 2.1 to 2.9, the preferred options all fall under the balanced approach (overview shown in the table below).

Objective	Status Quo	Supportive	Balanced	Prescriptive
1: Consumers are better able to identify and avoid or manage potentially harmful content	0	11	13	5
2: Consumers are incentivised to avoid potentially harmful content	0	12	13	2
3: Creators of content (including social media users) are better able to anticipate and avoid or minimise the potentially harmful impact of that content	0	9	10	1
4: Creators of content (including social media users) are incentivised to minimise the potentially harmful impact of that content	0	6	9	-1
5: Creators of content (including social media users) are incentivised not to create very harmful content	0	2	16	9
6: Publishers/providers of media and online content are mandated to minimise the prevalence and impact of potentially harmful content	0	2	16	4
7: Publishers/providers of media and online content are incentivised not to promote potentially harmful content	0	12	19	13
8: Publishers/providers of media and online content are incentivised not to publish/provide very harmful content	0	-4	14	5
9: Consumers are prevented from experiencing extremely harmful content	0	-2	8	4
Total	0	49	116	31

252. The balanced approach takes a balanced mix of Government regulatory intervention and a focus on partnering with industry and community stakeholders to achieve objectives and best address the policy problem. This would allow Government to have

extended regulatory oversight of content and be proactive about responding to content across all types of harms.

253. For six out of the nine objectives the balanced approach was the clear preferred option which were:

- Objective 4: Creators of content (including social media users) are incentivised to minimise the harmful impact of that content.
- Objective 5: Creators of content (including social media users) are incentivised not to create very harmful content.
- Objective 6: Publishers/providers of media and online content are mandated to minimise the prevalence and impact of potentially harmful content.
- Objective 7: Publishers/providers of media and online content are incentivised not to promote potentially harmful content.
- Objective 8: Publishers/providers of media and online content are incentivised not to publish/provide very harmful content.
- Objective 9: Consumers are prevented from experiencing extremely harmful content.

254. The supportive approach was very close to the balanced approach as the preferred option for three of the objectives which were:

- Objective 1: Consumers are better able to identify and avoid or manage potentially harmful media content.
- Objective 2: Consumers are incentivised to avoid potentially harmful media content.
- Objective 3: Creators of media content (including social media users) are better able to anticipate and avoid/minimise the harmful impact of the content.

The balanced approach is the most preferred option to address the policy problem

255. Overall, it is clear that the balanced approach would be the most preferred across all objectives with a total score of 117 compared to the supportive approach (47) and the prescriptive approach (32). While the supportive approach for objectives 1-3 came close as the preferred approach, the balanced option provides additional levers that makes it better able to achieve these objectives and as a result better address the policy problem. More detailed analysis about the balanced option and specific detail about what this would look like as a new content regulatory framework is to be determined following public consultation on the discussion document.

256. It should be noted, however, that the balanced approach anticipates the use of levers that are also envisaged under both the other options. This means that, within the overall framework of the balanced approach, the development of detailed proposals to address particular objectives may closely resemble the options attributed to the supportive or prescriptive approaches in the above analysis.

The balanced approach will enable further consideration of embedding Te Tiriti o Waitangi principles in the detailed design and implementation of the new framework

257. The balanced approach presents the best opportunity for Government to more effectively embed Te Tiriti principles within the governance arrangements and outcomes of the new framework. In addition, it would enable processes to embed greater community collaboration and decision-making. The levers available under this approach would allow Government to have regulatory oversight of content, while enabling it to be proactive about responding to all types of content-harms through

opportunities to partner with Māori as Treaty partners, communities, and Media Service Providers on how best to address content-related harms.

258. In particular, the balanced approach would empower Government to incorporate the principles of:

- **Kāwanatanga:** A balanced approach would reflect New Zealand’s social and cultural values, while also enabling consistency with the content regulatory initiatives of like-minded countries. This is particularly important for online content regulation, which has a strong international dimension;
- **Rangatiratanga and partnership:** Under the balanced approach, governance within the new regulatory framework would provide for Māori participation (e.g. Board membership, roles within regulatory oversight body), increased capacity and capability to understand how Māori engage with content and to ensure regulatory practice builds in tikanga, and education programmes and other initiatives that would be shaped by mātauranga Māori; and
- **Active protection:** Māori may experience harm differently or more frequently than New Zealanders generally, and the delivery of the new framework under a balanced approach would recognise this (e.g. through education levers, ensuring restoration processes where harm does occur are informed by tikanga and te ao Māori). The Crown must also protect the rights of Māori to freely express themselves and to generate and share content, especially to protect te reo me ōna tikanga.

What are the marginal costs and benefits of the option?

259. Note that, as the detail of the proposed approach has yet to be developed beyond the high-level description to be tested in the discussion document, detailed assessment of marginal costs and benefits is not possible at this stage. The table below provides an indication of whether these are expected to high, medium or low for each relevant group.

Affected groups <i>(identify)</i>	Comment <i>nature of cost or benefit (e.g., ongoing, one-off), evidence and assumption (e.g., compliance rates), risks.</i>	Impact <i>\$m present value where appropriate, for monetised impacts; high, medium or low for non-monetised impacts.</i>	Evidence Certainty <i>High, medium, or low, and explain reasoning in comment column.</i>
Additional costs of the preferred option compared to taking no action			
Regulated groups (MSPs/Industry)	All MSPs/Industry would be captured by the new regulatory framework to minimise harm of content	High	TBD
Regulated groups (User-content creators)	Some user-content creators would be captured	Medium	TBD
Regulators (Government)	Government to develop legislation for the new framework	Medium	TBD

Regulators (Operating costs of New Regulator)	New Regulator to be created under the new framework	High	TBD
Other (Educators/training bodies/NGOs)	New initiatives to be implemented and delivered	Medium	TBD
Other (Consumers)	Consumers have less access to problematic content	Low	TBD
Other (Society and Communities)	Some constraints on how identity is expressed	Low	TBD
Total monetised costs	Unknown	N/A	N/A
Non-monetised costs	Unknown	N/A	N/A
Additional benefits of the preferred option compared to taking no action			
Regulated groups (MSPs/Industry)	Increased consumer trust	Low	TBD
Regulated groups (User-content creators)	Clarity about acceptability of content	Low	TBD
Regulators (Government)	Improved social wellbeing	Medium	TBD
Regulators (New Regulator)	Clear and comprehensive mandated powers	High	TBD
Other (Educators/training bodies/NGOs)	Coordinated work programmes	Medium	TBD
Other (Consumers)	Less exposure to risk of harm	High	TBD
Other (Society and Communities)	Improved social cohesion	Medium	TBD
Total monetised benefits	Unknown	N/A	N/A
Non-monetised benefits	Unknown	N/A	N/A

Section 4: Delivering an option

How will the new arrangements be implemented?

260. Subject to confirmation and development of the preferred option following consultation, it is anticipated that implementation would be largely based on the enactment of a legislative framework prescribing core Government's responsibilities, establishing and empowering a single Regulator that will operate independently from core Government, and mandating MSP obligations.
261. The development and funding of a range of education/training programmes could sit outside the legislative framework although related roles for the Regulator and possibly other bodies could be authorised in the statute. As the options have been developed at a high level, detailed proposals for the implementation, timeframes, managing potential risks and undertaking any other work that may be necessary, will be developed and assessed following public consultation on the discussion document.

Legislative framework

262. New legislation is expected to underpin and coordinate most, if not all, roles in the new framework. The legislative framework will need to establish the legal status and accountability of the Regulator and specify its functions, powers and responsibilities. These include accountability and reporting requirements. Under the preferred option, the legislation would also require MSPs to adopt content harm minimisation codes and specify core requirements and standards for them. It would prescribe civil penalties and criminal offences, and sanctions for non-compliance with statutory requirements. New regulatory or enforcement tools, such as powers to block access or filter content would be included in legislation. In addition, new appeal and review processes would be needed to reflect the shift to a co-regulatory industry regulation approach. Principles of natural justice, legislative design guidelines and adhering to rights under the New Zealand Bill of Rights Act would guide further work to determine appropriate appeal and review pathways.

Core Government (Ministers and core departments) responsibilities

263. Responsibilities for system stewardship and governance would be central to core Government's function. But it would have very limited roles in the operation of the framework. These would be limited to:
- undertaking criminal prosecutions and civil proceedings at the request of the Regulator;
 - providing funding for public awareness campaigns; and
 - delivering necessary changes to the education curriculum.

Arms-length Regulator

264. The Regulator will need to be visibly at arms-length from executive Government. The final form of the Regulator will be determined following consultation and feedback on the functions and powers it should have. One option, among others, is that the Regulator may take shape as a Board group, appointed by government but with completely independent decision-making powers. It will have broad roles in relation to harm minimisation as well as specific functions, all mandated in the legislation. The broad roles would include publicity and public information, advising Government (and Parliament) on relevant issues, monitoring and research. Specific functions would include:
- providing guidance for, approving and monitoring Harm Minimisation Codes;
 - taking action to force the removal of extremely harmful content; and

- requesting and supporting prosecutions.

265. It could also have oversight over how censorship tools, such as take-down and service disruption measures, are designed, applied, and governed. Strong Māori representation would be built into the development of this regulatory Board, to ensure that Māori representation and influence is present during all stages of content regulation.

Media Service Provider (MSP) obligations

266. As a reminder, MSPs is the term we use for all entities who provide or enable access to content to the public. MSPs will have a general legislative obligation to prevent and minimise the risk of harm from their services, and a specific requirement to adopt and implement an approved Harm Minimisation Code. The code would be required to cover how harmful content is prevented or removed, how consumers can make complaints, how customers are alerted to potentially harmful content, and how the entity will report on these measures.

Harm Minimisation Codes

267. As noted at footnote 23, details of Harm Minimisation Codes including how they will be developed, monitored and enforced are set out in the discussion document for consultation. Broadly, codes may be developed at sector or industry-level or by bodies representing a sector or industry. Codes could also be developed for individual MSPs based on the specific services they provide and risk of harm from such services. Community and civil society input into the development and compliance of these codes would play an essential role. Sector/industry bodies may also have a role in administering complaints mechanisms and delivering industry training. Codes would need to be approved by the Regulator.

Education and training programmes

268. In addition to the above, there will be a variety of providers of education and training programmes. This will include reliance on the Education system to develop media awareness and critical thinking through its core curriculum content. It will include NGOs or other agencies contracted to provide public awareness type campaigns on behalf of the Regulator. It might also include industry training bodies or sector groups providing specific awareness and cultural competence training to journalists and other media sectors.

How will the new arrangements be monitored, evaluated, and reviewed?

269. As the options have been developed at a high level to date, details about how the new arrangements will be monitored, evaluated, and reviewed will be determined following the end of public consultation on the discussion document. However, as a key strength of the balanced approach is flexibility to adapt to changes in technology, content and social expectations, it is inevitable that monitoring and transparency reporting, and continuous review procedures, will be critical elements of the operation of the system. The separation between core Government stewardship of the overall system and the expected policy advice responsibilities of the arms-length Regulator will support ongoing monitoring, evaluation and adjustment of the system.

Appendix A: Additional information from the 2022 report ‘What we’re watching: New Zealanders’ views about what we see on screen and online’

270. The Classification Office undertook a nationwide online survey from 22 February to 14 March 2022, involving 1,201 New Zealanders. This included 1,001 adults and 200 young people aged 16 or 17. A combination of pre-survey quotas and post-survey weighting was used to ensure the results are representative of all New Zealanders by age within gender, region, ethnicity, and household income by household size.
271. The research report provides insight into New Zealanders’ views about the type of harms that come from content and the areas that cause most concern. The research project focused on the following areas:
- New Zealanders’ views about specific types of content and the potential for harm to children, young people and the wider community.
 - Perceptions of the influence or impact of this content.
 - People’s personal experience of potentially harmful content.
 - Ways in which people manage content to keep themselves and their families safe, and perceptions about age ratings and the classification system.
 - Views about how to mitigate potential harms of online content, including measures by industry and government regulation.
272. Key findings of the project include:
- 74% of respondents are concerned about children and young people seeing harmful or inappropriate content in movies, shows and video games;
 - 83% of those surveyed (aged 16 or over) were concerned about harmful or inappropriate content children and young people can find on social media, video-sharing, or other websites;
 - depictions of sexual violence or sexual harassment (95%) and self-harm or suicide (94%) are seen as especially harmful to children and young people, along with racist comments, behaviour or stereotypes (91%) and realistic violence (91%);
 - 97% believe that such content can have a negative influence on children and young peoples’ attitudes or emotional wellbeing;
 - 53% said they had seen online content in the past year that in their view promoted or encouraged certain attitudes or behaviours, such as hate or discrimination, terrorism or suicide; and
 - 42% ‘somewhat’ or ‘strongly’ agreed that it was hard to avoid harmful or offensive content online (27% disagreed).

Appendix B: International Approaches to Media and Online Content Regulation

Australia

273. Australia's Online Safety Act 2021, which took effect on 23 January 2022, sets out the Australian Government's expectations regarding the role of the online industry in ensuring online safety.
274. The Online Safety Act establishes a co-regulatory approach to harmful content on online platforms. This approach tasks industry organisations with developing industry codes regarding harmful online content, including pornography, which may be harmful to children but legally accessible to adults; R18+ material or material not restricted to people over 18 years of age access to which may require the establishment of a restricted access system; and content that can cause serious harm such as child sexual exploitation or terrorist content.
275. Online platforms are also tasked with developing reporting and complaints mechanisms regarding this harmful content, may be required to establish a restricted access system for R18+ material provided from Australia, and are required to notify the Commissioner of offending material reported on their platforms.
276. The Act also provides the eSafety Commissioner with added powers to regulate the industry, such as the establishment of industry standards in cases where the Commissioner finds industry codes to be ineffective or insufficient, which industry actors would be required to follow.
277. In addition, the Act introduces 'schemes' related to specific forms of online harm, including cyber-bullying material targeted at an Australian child, non-consensual sharing of intimate images, cyber-abuse material targeted at an Australian adult, and material depicting abhorrent violent conduct.
278. These schemes give the Commissioner the power to issue removal notices to online service providers, end-user notices to individuals, blocking notices to Internet Service Providers, and link deletion notices or app removal notices to search engines and app distribution services related to content, links, or applications that continue to contain or share harmful content.
279. The Act also grants the Commissioner with information gathering and investigative powers to investigate reports of harmful content. Reporting and review processes ensure the transparency and consistency of the Commissioner's actions.

United Kingdom

280. In 2019, the UK introduced tough new measures through an 'Online Harms White Paper', with the intent to make the UK the safest place in the world to be online. On 15 December 2020, the UK released its Full Government Response to the Online Harms White Paper (the Response paper) and has since announced an Online Safety Bill will be introduced in 2021.
281. The White Paper and ensuing Bill aim to tackle a range of harms including the incitement of violence and violent content, suicide, disinformation, cyber bullying and children accessing inappropriate material. There are also strict requirements for companies to take even tougher action to ensure they tackle terrorist and child sexual exploitation and abuse content.
282. The Response paper builds on the Online Harms White Paper published in April 2019, which proposed a comprehensive regulatory framework, addressing a spectrum of online harms from illegal to legal but harmful content. The Bill proposes a new regulatory system for online content, which creates a duty of care on online platform operators to protect their users. Platforms in scope are hosts of user-generated content

which can be accessed by users in the UK; and/or facilitate public or private online interaction between service users, one or more of whom is in the UK; and search engines.

283. The “duty of care” framework is aimed to be proportionate and risk-based, ensuring companies in scope have the appropriate systems and processes in place to respond to harmful content and activity, while protecting users’ rights. The regulatory framework will be set out and overseen by its existing communications regulator (Ofcom), five of which will be given broad powers, including the ability to enforce fines of up to £18 million or 10% annual turnover.
284. The regulatory framework will have a two-tiered approach, imposing additional requirements on the most high-risk, high-reach services (category 1), to the requirements of most services (category 2).

Ireland

285. In January 2022, Ireland’s Government approved the publication of the ‘Online Safety and Media Regulation Bill’. The Bill established the Media Commission as a new regulator and approved the dissolution of the existing regulator, the Broadcasting Authority of Ireland.
286. Ireland had significant regulatory and legal frameworks in place in relation to many online issues, including data protection and criminal justice responses to criminal activities online, prior to the introduction of the Online Safety and Media Regulation Bill. However, there was a gap both internationally and in Ireland when it came to addressing harmful online content. This new law closes the legal gap and establishes a robust regulatory framework to deal with the spread of harmful online content.
287. Through the Bill, The Media Commission was granted regulatory powers, from imposing levies, to blocking access to certain online content, issuing notices to end non-compliance, and prosecuting senior management for failing to comply.
288. The framework also provides a process for defining “harmful online content”, introduced the making of binding and non-binding online safety codes, provided a risk-based process for designating online services for regulation, and introduced a “super-complaints” scheme for systemic issues to be brought to the attention of the Media Commission.
289. Online safety codes will deal with a wide range of issues, including measures to be taken by online services to tackle the availability of harmful online content, for example cyberbullying material, on their services.

France

290. France’s content regulation policies and frameworks have raised significant, international, concerns relating to freedom of speech, and as such, their high-level principles and content review aims do not translate to those of New Zealand.
291. In France, what is illegal offline is now illegal online. This came following the introduction of new hate speech laws, known as the Avia law, on July 1, 2020. The law applies to hate speech, and material promoting terrorism or child abuse. These controversial and highly criticised regulations require social media companies to

remove certain illegal content within 24 hours of being flagged by users – and in some cases, as little as an hour; to avoid facing fines of up to €1.25 million.

292. The European Commission requested France hold off on passing the law until the Digital Services Act, an overhaul of how the EU regulates digital platforms, was rolled out across the European Union. However, French officials ignored these concerns.
293. France focuses specifically on disinformation around electoral campaigns. As a result of the electoral interference in 2017, France targets these threats to democracy by requiring platforms to meet certain standards of conduct during the three months preceding elections, and arming judges with the power to order content removal during election campaigns. Regardless, disinformation is not legislated as a threat, nor has it resulted in the creation of ministerial organisations.
294. 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. Paris has shown strong support for the Christchurch call since May 15th, 2019.

European Union

295. The European Union passed its terrorist content directive in April 2021, which requires internet companies to take down terrorist material within an hour of receiving a notice from an authorized government agency of an EU member state. The European Digital Services Act 2020 (DSA), imposes greater liability on online companies for illegal material on their systems, revising the 20-year-old Electronic Commerce Directive that established this online immunity in the first place.

Finland

296. Finland is known for its freedom of the press, approach to education around misinformation, and one of the world's first digitalised public services. Ranking first out of 35 countries on the Media Literacy Index and first in public trust in the news media, means Finnish citizens are less likely to turn to alternative sources for news.
297. Other countries are looking to Finland as an example of successful resistance to mis and disinformation, and 'fake news'. In 2019, the government ran a public campaign during the parliamentary elections encouraging citizens to be alert to foreign attempts at electoral interference.
298. Finland also places significant value on education in schools. In 2016, 'multi-platform information literacy' and 'strong critical thinking' became a core, cross-subject component of their national education curriculums.
299. The Prime Minister's Office ran the campaign "Finland has the best elections in the world – and why is that", which encouraged citizens to read the media critically, increase public awareness of where to find reliable official information about the elections, and how to recognise it and prepare for it.
300. School curriculum dedicated to building media literacy is incorporated into all subjects, making it a fundamental value of education. The curriculum is part of a strategy by the

Finnish government which responded to targeted fake news stories from Russia in 2014.

301. The Act on the Classification of audio-visual programmes provides guidelines for the classification of television programs for the protection of children against pornography and violence. Violations of the act are punishable by Finnish penal code.
302. The Act on the Exercise of Freedom of Expression in Mass Media specifies that anyone can start a periodical publication. Information about the publisher and editor must be available in the publication. The orientation and content of the publication is entirely for the responsible editor to decide – the same principles apply to online publication.
303. The principle of media freedom is enshrined in the Act on the Exercise of Freedom of Expression in Mass Media. The Act brings the press, broadcasting, and online media within the same framework of responsibility and freedom of expression. Anyone who has a justifiable reason for considering themselves a victim of an offence arising from content has the right to have a reply published in the same publication or programme.
304. Finnish authorities have also been discussing the possibilities of obliging online publishers and service providers to prevent unlawful content on the websites they administrate. They would be required to pre-moderate discussions and spontaneously delete inappropriate messages, and failure to do so would result in punishment. This would, however, go against Finland's traditionally uncensored approach to media. The government has also proposed that blog-owners must remove racist, child-pornographic, and terrorism-inciting messages from the discussions in their blogs. These plans have generated a lot of criticism among different advocacy groups.

Estonia

305. In 2007, a disinformation campaign escalated into a cyberattack on Estonian government websites, banks, and media outlets – thought to be the first attack of its kind on an entire country. In response, Estonia invested in online infrastructure to prevent future attacks. This has developed into the digital society of Estonia today (known as 'e-Estonia').
306. Estonia now considers itself to be the "world's most advanced digital society". Most of Estonian civic life is digital with public services such as healthcare, education, business, city infrastructure, e-Identity, and voting, provided digitally. Artificial Intelligence (AI) technology is also likely to increase as Estonia seeks to maintain its reputation as innovators in communications and information. These steps are likely to pose challenges for policymakers as the sector continues to develop.
307. Estonia ranks high in media freedom on the Media Literacy Index and is thought to be well equipped to deal with disinformation. Children in Estonian public schools, from kindergarten to high school, are taught media literacy. In elementary and middle school age (age 5-13), media literacy is incorporated into other subjects rather than being a standalone course. In high schools, a 'media and influence' course focusing on the role of media and journalism is mandatory for students. This investment in education is thought to facilitate a culture of, and strengthen the ability to, critically analyse information and messaging.
308. According to Facebook and Twitter's July-December 2020 transparency report, neither company received any content removal requests from the Estonian government.

Google, however, received four takedown notices from the government and removed 1.7% of the reported content.

309. Media services are regulated by the Media Services Act (MSA). Estonia also has various laws aimed at protecting children online, such as the Child Protection Act which limits exposure of certain content for children below the age of 18. This includes content that promotes violence or cruelty.
310. In February 2021, a Bill transposing the EU Audio Visual Media Services Directive (which requires platforms such as YouTube to take appropriate measures to protect minors and the public from harmful content) was implemented in Estonia. Implementation of the Bill faced significant delays due to a lack of consensus around definitions such as what constitutes the “incitement of hatred”. The definitional challenges experienced by Estonian policymakers could provide context for New Zealand’s Review and similar challenges that may arise around defining “harmful” content.
311. However, internet access and online content in Estonia has very few limitations. Instead, the focus is on protecting user rights and maintaining freedom of expression. Online illegal gambling websites are the most restricted, yet most political, social, and cultural content is freely available to users. Mis and disinformation are significant concerns for the Estonian government, so considerable emphasis is placed on maintaining trust in the state and democratic institutions, as well as building media literacy through education and awareness.