

23 February 2024

Ofcom Online Safety Team  
Ofcom  
Riverside House  
2A Southwark Bridge Road  
London SE1 9HA

By email: [IHconsultation@ofcom.org.uk](mailto:IHconsultation@ofcom.org.uk)

Tēnā koe,

## Re: Consultation on protecting people from illegal harms online

The New Zealand Classification Office—Te Mana Whakaatu welcomes the opportunity to provide feedback on the first phase of Ofcom’s draft online safety guidance. Thank you for inviting us to participate in this process, as observers to the Global Online Safety Regulators Network. We will soon follow this up with feedback on the second phase of consultation addressing protections to children.

The Office is a content regulator, but not a service regulator. We have focused our feedback on the areas where we have expertise or evidence to offer. This reflects our wealth of experience in classifying content, researching harms, educating the public and providing resources to empower New Zealanders to make informed choices about what they watch, and to protect themselves and their children and young people.

We are impressed with how sophisticated the proposals are. They comprehensively address many of the features we believe are necessary for effective, accessible, and fair content regulatory systems, both in Aotearoa New Zealand and globally.

New Zealand has not yet passed content regulation at the scale of the Online Safety Act, and we watch on with interest as the United Kingdom works to implement these changes.

This response is not confidential, but the views expressed here are those of the Office only.

### About us

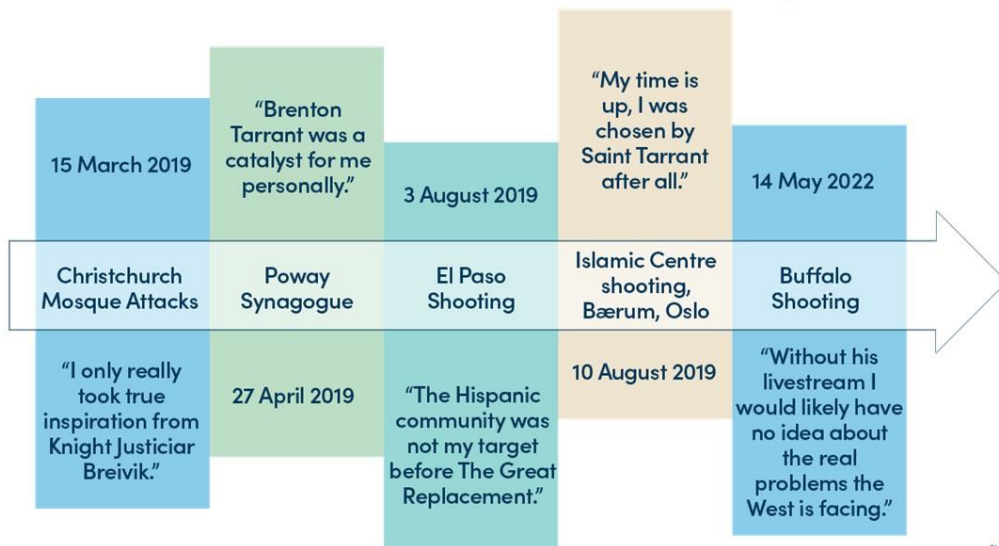
The Classification Office is an independent Crown entity in Aotearoa New Zealand. Our roles and functions are set out in the [Films, Videos, and Publications Classification Act 1993](#) (the Classification Act). Broadly, we:

- Classify physical content (such as films released in cinemas or on DVD) and material submitted by Crown agencies and the courts. The Chief Censor has statutory powers to restrict and ban some harmful content.
- Provide information, education, and resources to empower New Zealanders to make informed choices about what they, and their children and young people, watch.
- Support streaming services to rate their content for New Zealand viewers.
- Produce research and practical resources to help New Zealanders understand the classification system.
- Provide a complaints and enquiries service to the public.
- Maintain expertise in countering violent extremism to support the wider government response.

### Question 1

Our experience is that online harms are multiplicative. They can result from, and cause further, offline harms. This is starkly illustrated by the following diagram, which shows how the 2019 Christchurch mosque attacks were motivated by prior terror attacks, and motivated subsequent terror attacks, around the world:

## How events connect and inspire



We would be happy to send further information about the causal nature of some online harms, including literature on this topic.

## Question 2

### *What we know about content harm*

New Zealanders of all ages commonly see harmful content on screen and online. It can be difficult to avoid, and can impact negatively on wellbeing. When we survey the New Zealand public, most people say they are worried about the effects of harmful content, whether in movies, shows, games, and social media, or in other online spaces.

Our research shows that New Zealanders find it difficult to protect children and young people online, and that most people support regulating harmful online content. 53% of respondents to our [2022 survey](#) had seen online content that promotes or encourages harmful attitudes or behaviours (such as discrimination, terrorism, or suicide). 33% had seen content that directly promotes or encourages violence towards others, and 20% had seen online content that encourages some form of self-harming behaviour.

Evidence from New Zealand and overseas, and our own experience classifying content, tells us that certain types of content can cause serious harm to individuals and injure the public good.

This is a growing area of study with more research coming in. The US Surgeon General's recent [report](#) *Social Media and Youth Mental Health* outlines the indications that social media presents real risks to the mental health and wellbeing of children and young people.

### *Harm from content manifests in different ways*

Viewers can be disturbed or shocked by distressing material, suffering mental anguish and adverse psychological experiences. Viewers can be triggered to relive their own past trauma, so they rely on content warnings to make decisions about what to watch – when those warnings are available. Our recent [public survey](#) showed that most people think age ratings (79%) and content warnings (74%) are important when choosing a movie, show, or video game for children and young people.

Viewers may experience attitudinal harm from consuming content that depicts degrading, dehumanising, and demeaning conduct – including sexual conduct.

Some content can create or reinforce negative attitudes toward women and trans people, and perpetuate negative sexual stereotypes, or normalise extreme or unsafe sexual practices.

Content may encourage viewers to cause serious physical harm to themselves or others. They may also be encouraged to treat or regard themselves or others as degraded, dehumanised, or demeaned. Viewers may develop, normalise, or have harmful and antisocial attitudes reinforced, become desensitised to the effects of real-life violence or diminish their capacity for empathy and compassion.

Viewers may be encouraged to imitate content that glorifies risky, unsafe, or illegal behaviours – such as drug use and disordered eating. Adolescent girls in particular are at risk from content that perpetuates body dissatisfaction, disordered eating behaviours, social comparison, and low self-esteem. We aim to release a more comprehensive research resource on online misogyny and violent extremism in 2024.

### *Children and young people are especially vulnerable to harm*

Research into brain development from the [Collaborative Trust](#) shows that children and young people are disproportionately susceptible to harm because of their general levels of emotional and intellectual development and maturity. They have not yet developed the cognitive capacity to critically evaluate certain information. Exposure to age-inappropriate content can impair their mental, emotional, and social development.

[Children cannot always distinguish between what is real and what is not.](#) At 6–8–years-old, only 10% fully understand the difference. This increases to 36% by the time children reach their teenage years.

The [2020 Children’s Media Use: Research Report](#) by the New Zealand Broadcasting Standards Authority and NZ On Air found that:

- 87% of children have seen content on programmes and shows that has upset them. 72% of them have seen something online that has bothered them.
- Children found sex and nudity, violence/torture, and animal harm most upsetting. Parents have reported negative impacts on children’s behaviour: 20% had nightmares or difficulty sleeping, and 19% copied aggressive behaviours.

Netsafe's [Ngā taiohi matihiko o Aotearoa – New Zealand Kids Online](#) report found that:

- Almost 50% of young people have been exposed to potentially harmful online content. 28% of young people that were exposed to this sort of content said they were fairly or very upset by the experience. This emotional response was significantly higher for girls (38%) than boys (18%).
- 25% of 9 to 17 year-olds said that they had been bothered or upset by something that happened online in the last year. 46% of them said they were fairly or very upset by that online experience. This response was more common among girls and 12 to 17 year-olds.
- Nearly 20% of 13 to 19-year-olds experienced an unwanted digital communication (such as accidentally seeing inappropriate content online) that had a negative impact on their daily activities. 80% of those who reported experiencing an unwanted digital communication said they had an emotional response to it.
- 20% of teenagers had accessed self-harm material and some (17%) “how-to-suicide guides”. 15% had looked for information on “ways to be very thin”.

### *What our Youth Advisory Panel have said*

Since 2018, our office has engaged a Youth Advisory Panel (YAP) as part of our wider youth engagement strategy. The YAP is a diverse group of young New Zealanders aged 15 to 19, who provide input into our classification, research, and information work.

When we facilitated the YAP's engagement with domestic content regulatory proposals last year, they said:

#### **Young people want to control what they see on social media**

The Panel said that it was important for young people to make their own decisions about what they should or should not experience online and, in many cases, they are already trying to influence what they see. For example, they take care with what they actively 'like' on social media to shape algorithmic outcomes. They said that current complaints processes were inconsistent, and complaints needed to be dealt with faster.

#### **Young people want more information about social media practices**

The Panel felt that more transparency is needed on platforms, and information needs to be presented well and clearly, for example, to identify if a post has used a filter or has been photoshopped.

### **Young people feel that platforms have a responsibility to keep them safe**

The Panel felt that platforms need to take more preventative measures to keep users safe: “13-year-olds and 18-year-olds shouldn’t be shown the same content”.

### **Education, especially of older people, will be key**

Our YAP members said that online content can impact young people, and issues such as disordered eating can become a part of their offline reality. Education on the ways that platforms work and are used will be important to support not only young people, but also their parents: “Parents still don’t know how it all works”.

These concerns are consistent with [2023 research](#) with young people conducted by Te Hiringa Mahara—the New Zealand Mental Health and Wellbeing Commission. Their research found that:

... social media increases young peoples’ interaction with content that is not intentionally harmful but can cause distress because of the volume of information and the difficulty of shutting it off.

The young people interviewed for that research were:

... clear in identifying the responsibility of platforms in regulating what is published. They want to see more efforts to regulate material, protect young people from harmful messages and provide support for developing the skills and tools to understand what they see and hear online.

## **Question 4**

Consumers will benefit from a consistent experience across the various types of services that these measures have been proposed for. We think people should have similar experiences and expectations of content protections and warnings regardless of where that content is found or the form it is in. This applies not just to social media, search services, and other online content, but also to film, television, games, and other physical publications.

## Question 12

### *Detection and removal of content must be swift*

We endorse the comprehensive approach taken to developing the illegal content Codes of Practice. We emphasise that systems and processes for taking down illegal content must be designed to ensure swift removal.

We would also point out the benefits we have found in making fast quasi-judicial decisions about what is, and is not, illegal in New Zealand. Although United Kingdom law does not provide this function, regulators can support services to quickly identify content and apply legal tests, which is especially important in 'edge' cases. Risk-averse services may systematically remove legal content if they do not have adequate support, guidance and jurisprudence to know where 'edge' cases fall. These supports will help uphold freedom of expression.

Further to our response to question 26 (below), we think hash-matching will be beneficial not only for publications that are illegal, but also for those that have been found not to be.

### *Care will be needed around criteria for content removal, and associated metrics*

Performance targets for content moderation should not stray too far from the types of factors contained in the proposals. Targets should not include, for example, amounts of content removed, or amounts of referrals or requests actioned by the service.

It is also important that services comply with the Codes in ways that are grounded in evidence, and with reference to fair, evaluative criteria to determine the risk of physical, mental, emotional and reputational harm to individuals and communities, for example:

- depicted harm to victims (such as child sexual exploitation material)
- potential harm to victims (such as terrorist and violent extremist radicalisation content encouraging violence)
- potential harm to vulnerable people (such as instructional suicide sites).

## Question 21

Similar to our response to question 50 (below), information on the types of measures that public and private content will be subject to should be provided, and in a form that is accessible to the public, especially young people.

## Question 26

The Global Internet Forum to Counter Terrorism (GIFCT)'s hash-sharing database is a good example of how hash matching can result in fast and effective removal of terrorist content online amongst cooperating platforms. It is best used when judicial or quasi-judicial decisions like ours are the basis for hashing and take-downs, since the appropriate legal and human rights test have already been worked through.

## Question 28

### *A 'single front door' for complaints, reporting and requests*

Regarding Measure 2 ("All search and user-to-user services must provide an easy to find, easy to access and easy to use complaints system"), we would suggest considering a simple, centralised method for users to navigate services' complaints and reporting systems. Services may make complaints and reporting mechanisms easy to use on their own, but users must be able to get to the right mechanisms in the first place.

Regardless of whether Option 1 or Option 2 is adopted under this measure, the manner and extent of services' compliance with complaints system requirements will vary. Even high levels of compliance across the sector could lead users to be faced with a multitude of avenues through which they could, correctly or incorrectly, take the issues they have. These will include reporting content, appealing decisions, and requesting information.

A 'single front door' for complaints, content reporting and requests would avoid systemic fragmentation, and potentially motivate services to better comply with the Act. Such a solution should also allow services to innovate and implement complaints and reporting systems that are appropriate for them.

### *Cross-referencing/interoperability of supporting information for complaints*

When abuse occurs across multiple online services, victims can face the ordeal of having to establish grounds for their complaint on each service. This does not always provide a full picture of the scale of the abuse an individual has received. For the complainant, it can be exhausting and retraumatising to undertake substantially the same process multiple times over. For the system, this may result in inefficiencies, fragmentation and incomplete complaints responses.



Services could be required to make information supplied to them by complainants available to Ofcom and/or other services, with appropriate safeguards, to ensure that responses to online abuse are proportionate to victims' overall experiences, not just to what they have faced on one service.

### Question 33

Services will need to strike a balance between informing young users of the risk of possessing or sharing illegal content, against the risk that by informing them of this, they then decide to seek or share this content.

### Question 49

We endorse the proposed draft Illegal Content Judgements Guidance (ICJG).

Services should be required to have safeguards against false positives (such as over-removal of content, or over-referral to enforcement agencies), consistent with obligations to freedom of expression. This is especially important if services develop their own guides for moderating content based on the ICJG. Services will have the commercial prerogative to exercise caution and choose how they meet the regulatory requirements generally, however this may run up against their obligations to freedom of expression and corresponding expectations from consumers.

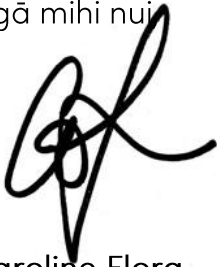
### Question 50

As well as making the ICJG accessible for services, it should also be provided in a form that is accessible to the public. Users must know, at least at a general level, which types of online content are illegal online, as well as the extent of measures that public and private communications will be subject to. This would ensure that the content regulatory system is fair and that users can seek information to avoid incrimination.

Ofcom could achieve this through its public education and outreach functions, along with an accessible summary of the ICJG.

We appreciate the opportunity to provide feedback on this consultation, and would welcome any further engagement with Ofcom as the implementation of the Online Safety Act progresses.

Ngā mihi nui

A handwritten signature in black ink, appearing to be 'CF', written in a cursive style.

Caroline Flora  
Chief Censor—Kairāhui Whakaaturanga Poumatua  
Classification Office—Te Mana Whakaatu