

Safer Online Services & Media Platforms

Document is here:

[https://www.dia.govt.nz/diawebsite.nsf/Files/online-content-regulation/\\$file/Safer-Online-Service-s-and-Media-Platforms-Discussion-Document-June-2023.pdf](https://www.dia.govt.nz/diawebsite.nsf/Files/online-content-regulation/$file/Safer-Online-Service-s-and-Media-Platforms-Discussion-Document-June-2023.pdf)

Submissions due: 31 July

Fact Aotearoa supports regulation to mitigate the harm of misinformation and the spread of calls to violence which persist in online spaces. We believe that a process of public engagement and discussion with platforms, backed by robust enforcement, can provide safety without stifling legitimate discussion. Indeed, without the thinly disguised threats of harm that currently abound, public debate will include more voices and allow more perspectives.

We urge supporters of Free Speech for all to submit in support of local regulation and reduce the control currently in the hands of the platform owners. We don't believe the owners are sufficiently aware of or committed to our local needs and concerns.

Please feel free to use our content to support your call for change.

You can just record that you agree that we need more regulation, that you agree with FACT Aotearoa's submission, or record your own responses.

FACT Aotearoa strongly supports this proposal. We feel that an appointment requiring the participation of our whole society in defining and monitoring harm is more democratic, responsive and based in local issues, and is superior to an appointment without these requirements, and to self-moderation by platforms.

We point out that opposition on the grounds of this being a political appointment leaves us with an unmanaged platform-based approach (except for the most harmful content). This is a less responsive, less democratic and less local process to define and monitor harm, so opposing on those grounds is disingenuous, as it prefers self-monitoring by unconcerned and profit-motivated platforms.

We applaud the requirement for education and want that well and urgently funded. We believe here is reason to suggest that projects building social cohesion and critical thinking skills should also be funded within certain parameters.

We have some reservations and concerns detailed below.

Definitions in the proposals

1. What do you think about the way we have defined unsafe and harmful content?

The definition should include actual harm or a credible incitement to harm irrespective of whether the content is experienced by those potentially harmed. Harm by direct engagement with harmful content is only one kind of harm. The proposal implies this, but is less explicit than we would like.

The definition should include risk if someone acted on the content, irrespective of the level of affective response in viewers. As an example, certain calls for execution are not currently deemed a sufficient threat, apparently, since they remain online, but calls for killing/execution/detention are certainly potentially harmful. Trivialising and gamifying actual harmful events should be recognised as a threat and included. We suggest

- Harm or threats implied or stated where actions would result in actual harm to people or property and they are literal threats, even if risk of action is low,
- Threats resulting from assertions or implications that past events are simulations or fabrications should be treated as harmful.

2. Does the way we have defined unsafe and harmful content accurately reflect your concerns and/or experiences relating to harmful content?

In terms of a description in law, there are limits to the definition of harm that can be used. The harm we see as presenting a threat that is more difficult to control is from: multitudinous small platforms amplifying misinformation producing an illusory truth effect (<https://thedecisionlab.com/biases/illusory-truth-effect>); strategies that create an impression of inflated numbers of supporters and attempt to present fringe opinions as normal; large funding support for products and boosting; significant amounts of high-quality resources created overseas and focussed on common themes in misinformation circles which are available here for misinformation spreaders¹; and the relentless attacks on institutions and individuals working against misinformation. Vaccine and health dis/misinformation is harmful and is proven to have deathly consequences on a large scale. There are some suggestions about such content below.

The EU includes advertising and on-selling in the equivalent legislation, and this and other possible ways of evading regulations should be considered

About our proposed new framework to regulate platforms

3. Have we got the right breakdown of roles and responsibilities between legislation, the regulator and industry?

No. The regulator should be empowered to demand standards at least as robust as those demanded by equivalent countries (eg the EU) around platform moderation, if there is public appetite for that. If the regulator is in a position to accept inadequate moderation when the public want better, we are in danger of being sold short.

We also suggest the regulator be given responsibility for content on smaller platforms, not as a monitor of these as we accept they are numerous, but in response to consumer alerts.

4. Do you agree that Government should set high-level safety objectives and minimum expectations that industry must meet through codes of practice?

Yes and they need to be funded, and outcomes robustly evaluated.

<https://www.axios.com/2021/02/23/memes-misinformation-coronavirus-56>

5. Do you agree with how we have defined 'platforms'? Do you think our definition is too narrow, or too broad? If so, why?

Relationships between platforms and subsidiaries needs to be clarified. Eg bookshops/publishers, blogs/host platforms, and advertisers. As these relationships are part of the way users find trustworthy information the widest net of associations possible should be regulated. In Q3 we suggest a second tier of harm mitigation, based on consumer alerts or complaints.

6. We are trying to focus on platforms with the greatest reach and potential to cause harm. Have we got the criteria for 'Regulated Platforms' right?

We feel that the abrupt disappearance of most regulation beyond the regulated platforms is problematic. Further to our response to Q3, we feel there is space for regulation beyond the most egregious or widely disseminated harms. Perhaps the regulator could be funded and empowered to deal with small-platform complaints and impose sanctions on small players. For example when the harmful content reaches a threshold of significance or reach through a single or multiple platforms, in balance. We think such powers could also interrupt the spread of warning-protected material from larger platforms, with the warning removed.

There could be some regulation of smaller entities like news platforms or advertising. News because it is deemed trustworthy and both because they offer easy evasion from regulation. We're already concerned that existing regulatory bodies for news and advertising are voluntary in nature and relatively weak in their powers.

The development of AI along with the identified and growing network of bot/private accounts boosting problematic material means that such content may become more engaging and varied, and it makes no sense to decline to legislate for this obvious threat. While enforcement may be challenging, the possibility of consequences will serve as a brake.

7. Do you think we have covered all core requirements needed for codes of practice?

The code should avoid opportunities to evade responsibility. For example the phrasing “and **aim to** align internationally, where possible” invites evasion (p35) We should require compliance if possible, not intentions or aims.

8. What types of codes and industry groupings do you think should be grouped together?

No comment on this question

9. Do you think some types of platforms should be looked at more closely, depending on the type of content they have?

Such scrutiny must be flexible as the online landscape changes quickly. Quarterly or annual review by the regulator might be required to identify and act on actual and anticipated changes. It makes sense to legislate as if we can achieve better monitoring of small platforms, for example, in case it becomes possible to do that.

10. Do you think the proposed code development process would be flexible enough to respond to different types of content and harm in the future? Is there something we're not thinking about?

The Regulator should have a requirement to plan for changing online landscapes.

11. What do you think about the different approaches we could take, including the supportive and prescriptive alternatives?

A prescriptive approach in consultation with platforms is reasonable. Platforms will be motivated to advocate for maximum allowable content as their profit margins are driven by content. Without enforceable authority to impact content there is little motivation to see sanctions as other than a content tax. With the sanction of reduced reach available, the regulator can negotiate from a position of strength.

12. Do you think that the proposed model of enforcing codes of practice would work?

A disruption of service option would be effective as a last resort. A regulation team rather than an individual regulator would give balance and authority.

13. Do you think the regulator would have sufficient powers to effectively oversee the framework? Why/why not?

The regulator must be sufficiently resourced and empowered to oversee the framework. If disruption of service is an available option, then the legislation is likely to be impactful. Overseas experiences of enforcement could guide this aspect of the legislation.

14. Do you agree that the regulator's enforcement powers should be limited to civil liability actions? (For example, issuing formal warnings and seeking civil penalties for non-compliance)

No. Temporary or permanent disruption of service should be a real, if rare, consequence. Overseas regulation gives insight into effective penalty limits. The effect of penalties on compliance should be regularly reviewed and penalty change or limit increase should be easy and straightforward.

15. How do you think the system should respond to persistent non-compliance?

Long term or permanent exclusion from provision in NZ. The precedent this would set globally is a powerful enforcer, as exclusion from an open democratic nation is a clear signal that some content is intolerable.

16. What are your views on transferring the current approach of determining illegal material into the new framework?

We have 2 reservations.
Age-related illegal content could end up unregulated in many spaces. We see no barriers to preserving such limits in face-to-face transactions and would like this obligation preserved.
We anticipate significant time between the establishment of a regulator and the finalisation of requirements and so current practice should remain until adoption of the standards.

About the regulator

17. Should the regulator have powers to undertake criminal prosecutions?

Yes. Local actors should be accountable for lawbreaking.

18. Is the regulator the appropriate body to exercise take-down powers?

Yes with appropriate requirements. They will be familiar with risks, impacts and overseas experiences.

19. Should takedown powers be extended to content that is illegal under other New Zealand laws? If so, how wide should this power be?

Absolutely. No illegal content should be accessible to ordinary users. The Regulator can play a part in this.
Take-down powers should be extended to other content, for example, content that requires removal due to rulings made under the Harmful Digital Communications Act, at present this can't be enforced by Netsafe and needs to be enforced by police requiring the user to remove it. The status means harmful content can remain online while police processes are followed, extending the amount of time harm can be caused.

20. If takedown powers are available for content that is illegal under other New Zealand laws, should an interim takedown be available in advance of a conviction, like an injunction?

Yes, with appropriate controls. Care should be taken that this is not used to, for example, stifle critique by mischievous complaints.
An example of a (possibly temporary) takedown for content not illegal could be when a case is filed under the Harmful Digital Communications Act, so that harmful content can be removed while pending a judgment. It is essential that any temporary takedown is evaluated promptly and this need must be part of the budget.
The regulator is well-placed to play a part in this decision which could be case-by-case.

21. What do you think about the proposed roles that different players would have in the new framework?

We think they are appropriate overall, but would like more requirements to include protected and vulnerable communities as decision-makers.

22. Have we identified all key actors with responsibilities within the framework? Are there any additional entities that should be included?

Protected and vulnerable communities should have a guaranteed voice.

What would the proposed model achieve?

23. What do you think about how we're proposing to provide for Te Tiriti o Waitangi through this mahi? Can you think of a more effective way of doing so?

Direct representation in decision-making would be better than "a voice"

24. Do you think that our proposals will sufficiently address harms experienced by Māori?

Māori should have a decision-making role and a minimum requirement for additional representation. Anti-Māori racism is often missed by overseas-based social media platforms, it is important to have culturally aware moderation of content, for the reason a requirement for large social media platforms to have Māori content moderators/consultants based in Aotearoa would be valuable.

25. What do you think about how rights and press freedoms are upheld under the proposed framework?

One of the greatest risks to press freedom comes from the spread of disinformation that has led to journalists facing threats of violence, as has been highlighted by research done at Massey University¹ and Reporters Without Borders.² Social Media platforms adhering to a code of practice will reduce the spread of disinformation and the incitement of violence against journalists, increasing press freedom. The restriction of journalist's access to the platform (as Twitter recently attempted), will also be regulated.

1

<https://www.stuff.co.nz/national/300584077/threat-to-our-democracy-kiwi-journalists-increasingly-face-violence-and-abuse-study-finds>

² <https://rsf.org/en/threats-and-violence-against-reporters-new-zealand-s-freedom-convoy-protests>

26. Do you think that our proposals sufficiently ensure a flexible approach? Can you think of other ways to balance certainty, consistency, and flexibility in the framework?

We'd like to see a mandated standard for consultation with civil society and the public.

Are there any other points you would like to raise?

An effective and safe transition to the new framework must not leave gaps in regulation while it is being constructed and made operational