Submission by

NZTech

to

Department of Internal Affairs

on

Safer Online Services and Media Platforms Discussion Document

31 August 2023

CONTACT:
Graeme Muller
Chief Executive
NZTech
E | Graeme.muller@nztech.org.nz **M |** +64 21 0252 0767

31 August 2023

NZTECH SUBMISSION ON SAFER ONLINE SERVICES AND MEDIA PLATFORMS DISCUSSION DOCUMENT

INTRODUCTION

1. NZTech welcomes the opportunity to comment on the Safer Online Services and Media Platforms discussion document (Discussion Document) and supports the Government's ongoing efforts to foster a safe internet environment for consumers.

2. While we welcome the Discussion Document's guiding principles, we recognise it will be complex to regulate traditional and online content under a single regime, and we recommend DIA continues to consult industry as this proposal evolves.

3. Our submission does not seek to address every aspect of the Discussion Document but instead focuses on what we consider to be key elements of the proposed framework. These include proposals for regulating content, determining services in scope of the proposed regulatory approach, and the proposed change in approach to the classifications process.

ABOUT NZTECH

4. NZTech is the peak body for the tech sector in New Zealand and a leading voice for the New Zealand technology ecosystem. We represent 20 tech associations with over 2,000 members who collectively employ more than 100,000 New Zealanders.

5. Our mission is to support a values-led, nationally connected tech community that is collectively lifting equity, sustainability and prosperity for all in Aotearoa New Zealand by creating jobs, export growth and impact through tech for good.

6. NZTech is a lead partner in the Digital Tech ITP, a collaboration between the New Zealand tech sector and the New Zealand Government, to help grow the sector.

7. NZTech is also the administrator for the Aotearoa New Zealand Code of Practice for Online Safety and Harms that has been adopted by Meta (Facebook and Instagram), Google (YouTube), TikTok, Twitch, and Twitter.

COMMENT

Creating safe online experiences

8. Like the New Zealand Government, NZTech and its members are committed to creating safe online experiences for New Zealander consumers.

9. In July 2022, NZTech and Netsafe launched an initiative that commits a range of technology companies to reducing the risk of online content that may cause harm to New Zealanders. The Aotearoa New Zealand Code of Practice for Online Safety and Harms has been adopted by Meta (Facebook and Instagram), Google (YouTube), TikTok, Twitch, and Twitter. The Code commits signatories to a set of Guiding Principles and Commitments that aims to mitigate the risks and reduce the prevalence of harmful content in seven areas: child sexual exploitation and abuse; bullying or harassment; hate speech; incitement of violence; violent or graphic content; misinformation; and disinformation.

10. NZTech commends the Government's leadership in establishing the Christchurch Call initiative. The Christchurch Call is a model for how multi-stakeholder initiatives can bring together governments, technology companies and civil society in an atmosphere of trust and to seek practical solutions to the challenge of terrorist and violent extremist material online. We have members who formally support the Christchurch Call initiative, and who have pledged to work with government and industry to address terrorism and violent extremist content.

11. Many of our members are also involved in other international efforts to tackle online child sexual abuse material, as well as initiatives to tackle terrorist and violent extremist content. These include membership and engagement with the Global Internet Forum to Counter Terrorism and the Technology Coalition and WePROTECT.

12. NZTech's members have extensive experience delivering services that prioritise customer trust and safety, and have navigated the development and introduction of content reform initiatives in other jurisdictions. NZTech welcomes the opportunity to facilitate ongoing engagement with our members, so that DIA can leverage their depth of experience as it undertakes this reform process.

13. Our members include those who operate content-based services, ranging from professionally produced content to content that is user-generated. These members have developed content policies, processes and resources designed to maintain safe online experiences, and which are proportionate and risk-based.

14. A risk-based approach to AI accountability is present in existing guidance such as the US National Institute of Standards and Technology's (NIST) AI Risk Management Framework. This framework has been developed through significant public consultation and "is intended to be flexible and to augment existing risk practices which should align with applicable laws, regulations, and norms". The risk-based approach is important for focusing regulatory efforts on high-risk AI without slowing down the pace of innovation around AI that poses little or no risk, e.g. AI being used to recommend fonts based on a document template. The EU's AI Act also takes a risk-based approach. A similar approach to harmonize a global framework for effectively combatting the potential harmful effects of AI has much to recommend it.

15. Some of our members have teams which work across content policy and enforcement, legal and law enforcement to combat harmful content on their services. These teams design and enforce content policies to meet customer expectations and to prevent harmful content. Depending on the service, these members may moderate for things such as nudity or images that are not appropriate for general audiences, offensive language, fraudulent behaviour, or terms of service violation (such as use of the service that violates the service's guidelines).

16. DIA and Government should incorporate digital content provenance and authenticity in guidance on responsible AI as effective tools for protecting consumers and creators alike. Specifically, platforms that generate content with AI should be required to include and preserve Content Credentials for content on their platforms. The transparency obligations for AI systems as originally outlined in the European Union's AI Act are important factors in bringing more trust to digital content.

17. Members may also deploy machine learning and automated detection tools, to scan for, screen and moderate content, in ways that preserve privacy. The combination of technology, tools and people ensures that content complies with our members' terms of service and stops the most harmful or violative content before it reaches consumers.

18. Some of our members offer a variety of services across the tech stack, and we observe that while many different technologies may play a role in making the internet and content on it available to users, not all services will be effective in content control mechanisms. For instance, mechanisms applied against some services may disable entire websites, apps, or even businesses and could affect customers who have not violated any laws. We recommend DIA adopts a "cascade" approach to responsibility for online content. The cascade approach means the regulator engages with the entity closest to the content, seeking to engage with subsequent downstream service providers only as a last resort and in a manner that understands the potential for over-removal of non-violating content.

Guiding principles

19. NZTech supports the Government's overarching objective in the Discussion Document, which is to ensure New Zealand's content regulatory framework is fit for purpose and allows for safe experiences on online services or media platforms for consumers. The Discussion Document's guiding principles articulate the Government's intention to introduce a new regulatory approach to reduce harms, including the establishment of a new regulator, where responses are proportionate, and focus on areas of highest risk.

20. These guiding principles acknowledge that responsibilities to ensure a safe and inclusive content environment are shared between individuals, platforms, and government. They also note that any intervention should be reasonable and be able to be demonstrably justified in a free and democratic society.

21. NZTech welcomes the guiding principles as we support an approach that balances harm prevention with freedom of expression, and which is based on a shared-responsibility model that enables proportionate responses to harmful content. Notwithstanding this, it will be complex to regulate traditional and online content under a single regime, and we recommend DIA continues to consult industry as this proposal evolves.

Approach to regulating content

22. The Discussion Document proposes bringing online and other media platforms into one cohesive framework with consistent safety standards, by creating codes of practice that set out specific safety obligations for larger or riskier platforms. NZTech supports this approach towards implementing a comprehensive self-regulating code of practice. Our own Code of Practice, which is well-established and widely accepted by members, addresses many of the issues raised in the Discussion Document. We believe such codes are effective ways of operating in the fast-paced and quickly evolving digital environment, and we would be happy to share with DIA our experience in developing and administering this code.

23. The Discussion Document seeks views on the definition of "unsafe" and "harmful" content. Harmful content is defined as where the "experience of content causes loss or damage to rights, property, or physical, social, emotional, and mental wellbeing", and unsafe content is defined as content "where there is a risk of harm occurring if that content was experienced by a person" and acknowledges that risk profiles differ between individuals.

24. As currently drafted, the definition of "harmful" is broad, and includes loss of rights, or loss or damage to property. These harms are dealt with through existing laws in New Zealand, and the examples of harmful content in the Discussion Document do not involve the loss of rights, or loss or damage to property. We recommend narrowing the

definition to exclude rights and property, to focus on harms more directly experienced through content which we believe is the intention of the reform proposals. For example, explicit references to the kinds of problems (such as eating disorders and violent material) envisioned by any future regulation or standards of practice would be advantageous for businesses.

25. We also recommend adding a materiality threshold, such as “significant harm” or “serious harm”, and establishing a causal link between the content and serious harm. This would help Regulated Platforms to identify such content.

26. We understand from the Discussion Document that the Government intends to focus on content that could cause the greatest harm. Regulated Platforms, which are those that meet certain criteria and will be subject to enforceable codes of practice, may be required to warn consumers or have policies or guidelines in place to limit or restrict access to specific categories of harmful content, with adult content in video games, and violent misogynistic content given as examples.

27. While we support specificity, adult content in video games, or violent misogynistic content, would arguably already be regulated through the classification system, or could be defined as “objectionable” material if the level of violence passed the threshold set under the Act. We recommend further consideration is given to whether this content is already adequately regulated before developing new enforceable codes addressing these specified categories of harmful content, as there is scope for duplication.

28. For clarity, NZTech understands that the Government does not intend to change the definition of “objectionable” or illegal content, which is defined by the *Films, Videos, and Publications Classification Act 1993* (the Act), and that legal but harmful content will not be subject to take-down powers. We support the continuation of this approach, as the category of content that is subject to a take-down request is clearly defined under the Act, and limits ambiguity and uncertainty for content service providers.

29. This legal clarity ensures that services can undertake targeted interventions and prioritise resources to restrict and remove child sexual abuse material, terrorist and violent extremist content, and other content as defined as “objectionable” under New Zealand law.

30. Civil penalties should take account of any good-faith effort by a company to comply. There may be valid cases for why a company does not take down content within a certain amount of time, e.g. it seeks clarification from the government or a reporter regarding the location of the content in question or it has questions around the validity of a takedown order. Ongoing exchanges with the government or the reporter should be treated as evidence indicating a good-faith effort to comply with the law.

31. Material that does not satisfy the definition of “objectionable” should not be subject to the take-down powers under the Act, as is currently the case. We understand the Government seeks to ensure community expectations are reflected in how regulated platforms moderate content, and desires to address a broader range of harmful content. We believe this is best achieved through a content service provider’s clear and transparent policies for the use of its platform by its users and audience.

Determining services in scope of the proposed regulatory approach

32. The Discussion Document defines a “platform” as “providers of content and services - for example, social media companies or broadcasters” that make content available to the public. It goes on to clarify that the definition excludes “platforms and services that exist primarily to enable other services and products”, with examples being websites of retailers, professional services, clubs and charities, as well as Internet Service Providers (ISPs).

33. In our view, the definition of “platform” is overly broad, and could inadvertently capture unintended targets. The definition and examples used for excluded targets are also confusing and are not sufficiently clear on the obligations applicable to the plethora of digital services available in New Zealand. Platforms that enable the sharing of content within one business or organisation, or are otherwise enterprise offerings (such as internal company websites or social apps, enterprise message boards) should be explicitly excluded.

34. As a starting point, we recommend that DIA clearly defines the intended targets of the proposed regulation. For instance, given that social media services have been referenced in the Discussion Document, DIA should draft a definition specific to that category of digital services. This could include acknowledging the role of engagement-based algorithms that amplify content and facilitate sharing in a manner that enables content to go “viral”, and functions that allow users to find and connect with others and “follow” or “subscribe” to content produced by particular accounts. This gives more legal certainty to digital service providers as to whether they would fall within scope. NZTech would be happy to work with DIA to propose draft language at the appropriate stage.

35. We also recommend that regulations affecting services that merely process, but do not control, content (such as cloud services, caching, Domain Name System services and ISPs) be carefully examined against the fundamental rights to freedom of information and expression. Where a service provider does not control what content is uploaded to the internet, how that content is made available to the public, and to whom it is made available, any potential order to remove or disable objectionable content risks the suppression of legal, protected expression. Cloud services could be an example as they have no control or knowledge of customer content on their systems and are processors rather than controllers of their customers’ data. They do not access or use their customers’ data other than as necessary to maintain or provide the services.

36. This important limitation means that they do not have the technical ability to perform the obligations contemplated for “Regulated Platforms”, such as taking down or disabling access to discrete pieces of content pursuant to content removal notices. They only have disproportionate options such as disabling the entire service. Cloud service providers should therefore only be engaged after outreach to the content provider has failed or the content provider has proven unwilling to comply, and DIA should understand that removal of only the objectionable content is highly unlikely.

37. Similarly, app stores do not control the content within apps. While app stores can apply age restrictions set by the app manufacturer, they cannot remove specific content from within the app, assess the age of the user of the app, or assess the age of the app’s intended audience.

38. Notwithstanding the additional clarity that is needed to understand how “platforms” will be treated, we understand that most obligations in the proposed framework would relate to “Regulated Platforms”, which are defined as “platforms where their primary purpose is to make content available” and exclude platforms and services that enable other services and products such as retailers’ websites, professional services and charities.

39. We support these exclusions and distinguishing between primary and ancillary services. This ensures that regulation is effectively targeted to capture the intended service type in a proportionate manner. Services should only be in scope if the activity in question is their “primary purpose”, rather than a small/secondary “ancillary purpose” of the service. Clear scoping ensures regulations are targeted to capture only intended service types in a proportionate manner.

40. In addition, as “Regulated Platforms” will be subject to greater compliance obligations, we recommend defining this group more narrowly to avoid capturing an unworkable swathe of general-purpose internet services. Consideration should be given to services that pose the greatest risk, such as social media services that contain user-generated content features that have the propensity to go viral and for content to be amplified, and which justify government and industry focus.

41. As noted above, most compliance obligations will be placed on Regulated Platforms. The first limb to determine whether a service is within scope of being classified as a Regulated Platform is whether the primary purpose of the service is to make content available. The second limb that DIA has proposed is whether the service has an expected audience of 100,000 or more annually, or 25,000 account holders annually in New Zealand.

42. The reach of a service is a factor for the regulator to consider, but it must be balanced with an assessment of whether a service presents a risk that should be subject to greater regulation. Large services should not be assumed to be inherently risky, as new and non-mainstream services can facilitate the proliferation of harmful content. A

service's characteristics pose different risk levels for different content types – for example, child sexual abuse material thrives in closed environments, while terrorist content or misinformation thrives in open, viral environments.

43. The proposed audience and account holder thresholds in the Discussion Document appear to be arbitrary, and do not contain an accompanying explanation on how these figures were determined. Should the reach of a service be a factor, we recommend raising the audience and account holder numbers, as smaller services could be overly burdened and would need to invest in a substantial technical uplift should they be classified as a Regulated Platform. They should only be required to do so if they represent a high risk for content that would be regulated under the new framework.

44. Consideration should be given to a more specific definition of Platforms, similar to the current definition of Regulated Platforms: "services where their primary purpose is to enable the creation and/or sharing of content publicly".

45. We further recommend balancing the second limb with an additional third limb, which dives deeper into the purpose or nature of the content on the service, which will determine its level of risk. Lower-risk services should be subject to lesser obligations, regardless of whether they satisfy the audience-reach threshold. For example, a general-purpose service serving a general audience, that primarily makes available professionally produced material to end-users, with limited ability to post content that has the propensity to go viral and be amplified, presents an inherently lower risk for online safety than a service that enables end-users to post or access high-impact materials such as adult content.

46. An arbitrary designation of a service may not result in intended outcomes for both consumers and services. A more nuanced approach to assessing whether a service should be classified as a Regulated Platform will result in better policy and safety outcomes.

A new approach to classification

47. The Discussion Document proposes to absorb the Classification Office into the new regulator and to cease the legally enforceable classifications system. On the latter proposal, we understand that age ratings would become recommendations through the code system, with platforms taking full responsibility for consumer advice.

48. This system is feasible for content such as films or television programs, where a service has a catalogue of content it provides to consumers and has a level of control over the content. Developing a self-rating system is complex and requires significant engagement with the relevant authority, as well as high-judgment decisions and understanding of context. In light of this, we recommend further details are provided on how this new system would operate under the proposed new regulatory framework. The treatment of user-generated content under the new system is unclear and should

be treated separately to a classification system given the potential impact on freedom of expression.

49. At this stage, it is difficult to provide further comments on the new approach to classification, and the workability across a varied range of content and their source, and we would welcome further engagement with DIA on this topic.

Regulator monitoring and enforcement

50. Proactive submission of periodic audits is not necessary. If a regulator discovers a problem with a regulated platform, DIA should notify the regulated platform of the problem and provide it time, if required, to furnish documents. We believe quarterly transparency reporting is too frequent. Industry practice is coalescing around semi-annual reporting. A supportive approach to enforcement makes the most sense. That way, the regulator can actually focus on the companies that present the highest risk of harm to New Zealand, as opposed to inhibiting business.

CONCLUSION

51. In a fast-moving digital environment, we believe self-regulating codes of practice as proposed by the Discussion Document would be effective in helping platforms provide safe experiences for their users. NZTech has a well-established Code of Practice which we would be happy to discuss further with DIA towards this goal.

52. Thank you for the opportunity to provide feedback on the Discussion Document.

Yours sincerely,

Graeme Muller

Chief Executive

NZTech

E| Graeme.muller@nztech.org.nz **P|** +64 21 0252 0767