Safer Online Services and Media Platforms – Submission

28 July, 2023

Authors: Dr John Kerr, Adele Broadbent, Prof Michael Baker

We thank the Department of Internal Affairs for the opportunity to provide a submission on the Safer Online Services and Media Platforms Discussion Document.

The focus of the PHCC is public health. We concentrate our submission in this area. While related to public health, we do not comment on some of the wider issues raised in the discussion document around hate speech, misogynistic and racist content, terrorism, and illegal content. These issues are important, but we will keep the scope of our feedback to a narrower set of issues.

We agree a new system is needed. We offer three specific points for consideration.

- **Acknowledging public health misinformation as harmful content**
- **Acknowledging promotion of certain products as harmful content**
- **Ensuring adequate access to platform content data for researchers**

Our first two points relate to Questions 1 and 2 in the discussion document:

> *1. What do you think about the way we have defined unsafe and harmful content?*
> *2. Does the way we have defined unsafe and harmful content accurately reflect your concerns and/or experiences relating to harmful content?*

1

Acknowledging public health misinformation as harmful content

We note that in defining unsafe content, the Discussion Document recognises that content can have wider impacts beyond the individual:

> *Some forms of content can be unsafe for communities, or for society as a whole – for example, if it discriminates against an ethnic group, or interferes in democratic processes. These forms of harm also fall within our proposals. (p. 18)*

We believe this definition is appropriate but recommend that there is more explicit recognition of false and misleading content falling within this definition, and in the Code created by the Regulator.

It is noteworthy that despite the general acknowledgement of the wider societal impacts of online content, the Discussion Document make no mention of misinformation, disinformation or false, inaccurate, or misleading information.

We acknowledge the definitional challenges around this kind of content and that the use of these terms varies. For simplicity we will refer to the general phenomena as **misinformation** in this submission.

Public health misinformation is an issue of national concern

The absence of any reference to misinformation in the Discussion Document contrasts with the DIA's *Summary of Initial Targeted Engagement* document, in which many of the groups consulted expressed concern over misinformation and conspiracy theories promoted on online platforms. Research commissioned by the DIA as part of the review also highlights misinformation as an *explicitly defined type of harmful content* and discusses its impact and possible regulatory approaches (Lips & Eppel, 2022; Thompson & Daubs, 2022).

Recent representative surveys in New Zealand have highlighted that New Zealanders are extremely concerned about misinformation. For example, a survey commissioned by the Classification Office found that an estimated 82% of New Zealanders are somewhat or very concerned about misinformation and 90% believe that misinformation is impacting public health in New Zealand. That statistic bears repeating: nine out of ten New Zealanders

believe misinformation is impacting public health (Classification Office, 2021; see also, DPMC, 2023).

There is evidence of public health disinformation (ie, deliberately shared false information) disseminated via online platforms in New Zealand by foreign powers. The anti-vaccine news story shown below was promoted by Russian disinformation agents via social media in New Zealand. It was one of the top five pieces of promoted propaganda in New Zealand in late 2021, identified by Microsoft in a report on disinformation activities globally (Microsoft, 2022).

High Recorded Mortality in Countries Categorized as "Covid-19 Vaccine Champions". The Vaccinated Suffer from Increased Risk of Mortality

Data on Mortality in the Most Vaccinated Countries. Increased Hospitalizations of the Vaccinated

By Gérard Delépine
Global Research, April 17, 2022

Translate Website

Share 3014   Like 3K   Tweet 400   Email 365   Share 4395

First published on September 30, 2021

Since the beginning of the health crisis, the French government has claimed that early treatment was ineffective. It has imposed major restrictions on our freedoms, in particular on doctors' prescriptions.[1]

It has also promised that vaccination would achieve collective immunity, the end of the crisis and a return to normal life.

But the failure for 18 months of this so-called "health strategy" based on false simulations, innumerable lies, promises never kept, as well as the propaganda and fear campaign has become unbearable.

*In turn this has been followed by **the extortion of consent** to be vaccinated, by outright blackmail, while curtailing our freedoms to move and socialize, our right to work and engage in leisure activities.*

Are the current vaccines that they want to impose on us effective?

Such disinformation has been flagged as a key national security threat by the Department of Prime Minister and Cabinet, which notes the critical role of social media platforms in helping to mitigate the risks posed by widely distributed false information (DPMC, 2023).

Misinformation is also a Tiriti o Waitangi issue. In the targeted consultation for this review, Māori focus groups expressed concern over misinformation in their communities and the negative consequences. Indeed the *Summary of initial targeted engagement* states:

> *"Misinformation and conspiracies online have considerable potential to influence Māori, who already have historic reasons for mistrusting government and public initiatives."* (p10)

Actors responsible for sharing misinformation are deliberately targeting Māori through social media, leveraging this lack of institutional trust (Ngata, 2020).

Health misinformation has negative individual and social impacts

There is clear evidence that false and misleading information on social media platforms can negatively affect individual health choices, and that this effect can have wider societal impacts.

People act on misinformation in ways that are immediately harmful to themselves. For example, there have been numerous reports of New Zealanders taking potentially toxic substances promoted as a cure for, or protection from the Covid-19 virus. Documented New Zealand examples include forms of bleach, alcohol, and ivermectin (Cook & Brooke, 2021).

However the most prominent examples of the impact of health misinformation relate to vaccination. Misinformation can dissuade people from vaccination. In experiments where people are randomly presented with different kinds of vaccine information, those who read false information about vaccines are less likely to report intending to get vaccinated (Loomba et al., 2021). Notably this effect has been replicated in a randomised online experiment with New Zealanders, using examples of misinformation that were widely shared on New Zealand social media (Thaker & Subramanian, 2021).

Vaccine refusal has immediate consequences for individual health, leaving individuals more susceptible to infectious diseases. However, it also has wider impacts due to limiting herd immunity--the more people who are vaccinated in a population the lower the chances of an infectious disease being transmitted through the population. When people who can receive a vaccine choose not to, this means people who cannot be vaccinated (eg those who are immuno-compromised) are at greater risk of being infected. In addition, unvaccinated people have a higher risk of serious illness than vaccinated, so are more likely to be admitted to hospital, straining scarce resources, and further reducing healthcare capacity.

Beyond vaccination, health misinformation in New Zealand has led to attacks on infrastructure (such as 5G cell towers (Clent, 2020)), and public opposition and aggressive lobbying against evidence-based public health interventions (such as community water fluoridation (Broadbent et al., 2015)). We could list many more examples of negative impact.

Finally, at a broad level, sustained promotion of misinformation and conspiracies online can undermine institutional trust in the organisations which seek to improve public health. This effect includes undermining trust in the medical system (Gluckman et al., 2021; Jolley et al., 2022).

Online platforms can effectively limit the availability and spread of health misinformation

The COVID-19 pandemic highlighted not only the issue of misinformation, but also the capacity of online platforms to respond effectively to it.

Mechanisms employed by platforms during the pandemic included removal of material, downgrading visibility, content warnings and displaying countering evidence-based content alongside misinformation. However, much of this occurred only in response to public and political pressure (Baker et al., 2020; Cotter et al., 2022).

We encourage the proposed Code to include a provision for the Regulator to make recommendations to platforms in response to specific health misinformation issues as they arise.

We note that that similar concerns were recently raised in the UK House of Lords debate on the Online Safety Bill. Baroness Merron's amendments in particular seek to place a duty on platforms to assess and report on health misinformation, defined as false or misleading information which "presents a material risk of significant harm to the health of an appreciable number of persons." (UK House of Lords, 2023, p. 18)

In advance of the Lord's debate, the organisation Full Fact sent a letter to Chloe Smith MP, UK Secretary of State for Science, Innovation and Technology (Full Fact, 2023). Signatories, including leading health organisations and experts, agreed that:

> *"One key way that we can protect the future of our healthcare system is to ensure that internet companies have clear policies on how they identify the harmful health misinformation that appears on their platforms, as well as consistent approaches in dealing with it. This will give users increased protections from harm and improve the information environment and trust in the public institutions."*

We reiterate these concerns in the context of the current New Zealand review.

Acknowledging promotion of certain products as harmful content

Another form of harmful content relates to the promotion of harmful products and services, particularly to young people. This includes promotion of tobacco, vaping, alcohol, and gambling.

Collectively these industries have an immense negative impact on the health and wellbeing of New Zealanders (Crossin et al., 2023; Ministry of Health, 2022).

These industries are restricted in their ability to directly advertise their products and services on mainstream media channels and on some social media platforms. However they have increasingly turned to 'user-generated' content as a channel for promotion. Rather than posting promotional content directly on social media channels, retailers encourage other users to post about their products. The use of affiliates and influencers to promote such products and services is widespread.

Examples include New Zealand vape retailers leveraging influencer accounts with large followings, and encouraging users to promote their brand by 'tagging' them in Instagram posts (Hardie et al., 2023). International research highlights similar strategies used by gambling and alcohol companies, though New Zealand-specific research is limited (Houghton et al., 2019; Thompson et al., 2023).
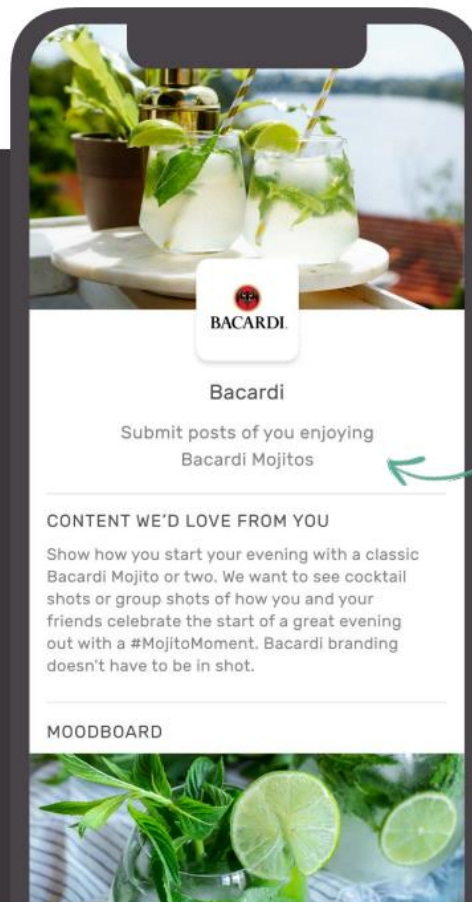
The most compelling evidence of negative impact from such promotion strategies comes from those responsible. For example, the PR firm Tribe promotes their services with a case study titled: "How UCG [User-Generated Content] helped generate a 14% uplift in sales of Bacardi" (Tribe Group, 2023).

The screenshot below highlights the firm's use of influencer user-generated content to create a buzz around the #MojitoMoments campaign. This example illustrates the methods by which companies can surreptitiously promote harmful products on social media, as well as their considerable impact.

Teaming up with TRIBE, OMD Create and Bacardi launched the

#MojitoMoment campaign,

aiming to use TRIBE's pool of 80,000+ diverse micro-influencers and creators to source an unprecedented amount of high quality user-generated content.

THE CTA WAS SIMPLE

BACARDI.

Bacardi

Submit posts of you enjoying Bacardi Mojitos

CONTENT WE'D LOVE FROM YOU

Show how you start your evening with a classic Bacardi Mojito or two. We want to see cocktail shots or group shots of how you and your friends celebrate the start of a great evening out with a #MojitoMoment. Bacardi branding doesn't have to be in shot.

MOODBOARD

User-generated content is typically not subject to advertising regulations.

The PHCC would like to highlight this form of harmful online content and recommend that the review consider how such content might be addressed by the Regulator.

There are existing examples of platform-self regulation that could be incorporated into the proposed Code, including limitations on user-generated and sponsored content promoting certain products (Kong et al., 2022).

Data availability for research on harmful content

The Regulator will have the capacity to commission research (Discussion Document Section 117, p. 60). Such research will inform much of the work around identifying and tracking harmful content. However large social media platforms have started to limit the ability of researchers to draw useful content data from social media platforms. In the last year, several major Platforms have reduced data availability through closing or restricting access to Application Programming Interfaces (APIs) or other interfaces allowing researchers access to machine-readable content and metadata.

Platforms must strive to be transparent and provide reasonable access to content, in a usable format, for researchers. We don't have a precise outline of how this requirement would look in practice but note that existing channels for researchers to access data on major social media platforms are heading in the opposite direction. Facebook/Meta (Albert, 2023), Twitter (Kupferschmidt, 2023), and Reddit (CITR, 2023) have all reduced access to content at scale via APIs or put in place prohibitive financial barriers for researchers using data for non-profit means.

We advocate for the Regulator to ensure provision of efficient access to platform content (through API or similar) to allow analysis of content, or a random sampling of content.

Access would not need to be fully public—though we would encourage this—but could instead allow access to vetted independent researchers, for research purposes. This approach would provide a critical way for independent organisations and commissioned researchers to evaluate the prevalence, spread and impact of potentially harmful content in New Zealand and more generally (see: Pasquetto et al., 2020).

The Regulator should not rely on aggregated reporting by Platforms. Analyses of such reports have found they are inconsistent and do not adhere to best practice (Urman & Makhortykh, 2023). Further, such aggregate reporting will not provide the regulator with insight into the types of harmful content addressed. Independent research can provide a better and more reliable insight into the availability and spread of harmful content and can inform efforts to mitigate offline impacts. However this can only occur if content and data are made available for analysis at scale.

Authors

Dr John Kerr is a Senior Research Fellow in the Department of Public Health, University of Otago, Wellington, and Science Lead for the Public Health Communication Centre. He has a background in science communication and psychology and a research focus on public health interventions, misinformation, and conspiracy theories.

Adele Broadbent is the Communication Lead for the Public Health Communication Centre. She is a former RNZ senior current affairs journalist with extensive communications experience in the health and international development sectors.

Prof Michael Baker is Director of the Public Health Communication Centre. He is a public health physician and Professor of Public Health as well as an award-winning science communicator. His research has focused on infectious diseases, environmental health, and housing and health.

References

Albert, J. (2023). *Facebook's gutting of CrowdTangle: a step backward for platform transparency*. https://algorithmwatch.org/en/crowdtangle-platform-transparency/

Baker, S. A., Wade, M., & Walsh, M. J. (2020). The challenges of responding to misinformation during a pandemic: content moderation and the limitations of the concept of harm. *Media International Australia*, *177*(1), 103-107. https://doi.org/10.1177/1329878x20951301

Broadbent, J. M., Thomson, W. M., Ramrakha, S., Moffitt, T. E., Zeng, J., Foster Page, L. A., & Poulton, R. (2015). Community Water Fluoridation and Intelligence: Prospective Study in New Zealand. *American Journal of Public Health*, *105*(1), 72-76. https://doi.org/10.2105/ajph.2013.301857

CITR. (2023). *Restricting Reddit Data Access Threatens Online Safety & Public-Interest Research (Open Letter)*. Coalition for Independent Technology Research. https://independenttechresearch.org/reddit-data-access-letter/

Classification Office. (2021). *The Edge of the Infodemic: Challenging Misinformation in Aotearoa*. https://www.classificationoffice.govt.nz/resources/research/the-edge-of-the-infodemic/

Clent, D. (2020). *Why setting cell phone towers on fire won't stop coronavirus* Stuff.co.nz. https://www.stuff.co.nz/national/crime/121537885/why-setting-cell-phone-towers-on-fire-wont-stop-coronavirus

Cook, M. A., & Brooke, N. (2021). Event-Based Surveillance of Poisonings and Potentially Hazardous Exposures over 12 Months of the COVID-19 Pandemic. *International Journal of Environmental Research and Public Health*, *18*(21), 11133. https://doi.org/10.3390/ijerph182111133

Cotter, K., DeCook, J. R., & Kanthawala, S. (2022). Fact-Checking the Crisis: COVID-19, Infodemics, and the Platformization of Truth. *Social Media + Society*, *8*(1), 20563051211069048. https://doi.org/10.1177/20563051211069048

Crossin, R., Cleland, L., Wilkins, C., Rychert, M., Adamson, S., Potiki, T., Pomerleau, A. C., Macdonald, B., Faletanoai, D., Hutton, F., Noller, G., Lambie, I., Sheridan, J. L., George, J., Mercier, K., Maynard, K., Leonard, L., Walsh, P., Ponton, R., . . . Boden, J. (2023). The New Zealand drug harms ranking study: A multi-criteria decision analysis. *Journal of Psychopharmacology*. https://doi.org/10.1177/02698811231182012

DPMC. (2023). *Let's talk about our national security: National Security Long-term Insights Briefing*. Department of the Prime Minister and Cabinet. https://www.dpmc.govt.nz/our-programmes/national-security/national-security-long-term-insights-briefing

Full Fact. (2023). Fact checkers and leading health organisations urge Government to take health misinformation seriously in Online Safety Bill. https://tinyurl.com/e35fbb9f

Gluckman, P., Bardsley, A., Spoonley, P., Royal, C., Simon-Kumar, N., & Chen, A. (2021). *Sustaining Aotearoa New Zealand as a Cohesive Society*. https://tinyurl.com/mpp26v59

Hardie, L., Mccool, J., & Freeman, B. (2023). E-Cigarette Retailers' Use of Instagram in New Zealand: A Content Analysis. *International Journal of Environmental Research and Public Health*, *20*(3), 1897. https://doi.org/10.3390/ijerph20031897

Houghton, S., Mcneil, A., Hogg, M., & Moss, M. (2019). Comparing the Twitter posting of British gambling operators and gambling affiliates: a summative content analysis. *International Gambling Studies*, *19*(2), 312-326. https://doi.org/10.1080/14459795.2018.1561923

Jolley, D., Marques, M. D., & Cookson, D. (2022). Shining a spotlight on the dangerous consequences of conspiracy theories. *Current Opinion in Psychology*, *47*, 101363. https://doi.org/https://doi.org/10.1016/j.copsyc.2022.101363

Kong, G., Laestadius, L., Vassey, J., Majmundar, A., Stroup, A. M., Meissner, H. I., Ben Taleb, Z., Cruz, T. B., Emery, S. L., & Romer, D. (2022). Tobacco promotion restriction policies on social media. *Tobacco Control*, tobaccocontrol-. https://doi.org/10.1136/tc-2022-057348

Kupferschmidt, K. (2023). Twitter's threat to curtail free data access angers scientists. *Science*, *379*(6633), 624-625. https://doi.org/10.1126/science.adh1696

Lips, M., & Eppel, E. (2022). *Mapping Media Content Harms*. https://tinyurl.com/y8et34v4

Loomba, S., De Figueiredo, A., Piatek, S. J., De Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, *5*(3), 337-348. https://doi.org/10.1038/s41562-021-01056-1

Microsoft. (2022). *Defending Ukraine: Early Lessons from the Cyber War*. https://tinyurl.com/mu8hmxns

Ministry of Health. (2022). Strategy to Prevent and Minimise Gambling Harm 2022/23 to 2024/25. https://www.beehive.govt.nz/sites/default/files/2022-06/Strategy-to-Prevent-and-Minimise-Gambling-Harm.PDF

Ngata, T. (2020). *The rise of Māori MAGA*. E-Tangata. https://e-tangata.co.nz/comment-and-analysis/the-rise-of-maori-maga/

Pasquetto, I., Swire-Thompson, B., & Amazeen, M. A. (2020). Tackling misinformation: What researchers could do with social media data. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-49

Thaker, J., & Subramanian, A. (2021). Exposure to COVID-19 vaccine hesitancy is as impactful as vaccine misinformation in inducing a decline in vaccination intentions in New Zealand: results from pre-post between-groups randomized block experiment. *Frontiers in Communication*, *6*, 721982. https://doi.org/10.3389/fcomm.2021.721982

Thompson, K., Dewolf, D., Macdonald-Spracklin, R., Frank, D., & Cooper, S. (2023). #Cheers: An Examination of User-Generated Alcohol Advertising on Instagram. *Substance Use & Misuse*, *58*(2), 204-210. https://doi.org/10.1080/10826084.2022.2151826

Thompson, P. A., & Daubs, M. S. (2022). *International Regulatory Frameworks for Online Content* https://tinyurl.com/5n89rytw

Tribe Group. (2023). *Bacardi's #MojitoMoment: A UGC Triumph*. https://www.tribegroup.co/stories/bacardi

UK House of Lords. (2023). *Online Safety Bill (HL Bill 87(Rev)—I)*. https://bills.parliament.uk/publications/50709/documents/3298

Urman, A., & Makhortykh, M. (2023). How transparent are transparency reports? Comparative analysis of transparency reporting across online platforms. *Telecommunications Policy*, *47*(3), 102477. https://doi.org/https://doi.org/10.1016/j.telpol.2022.102477