Policy Submission

Response to the AHRC's Discussion Paper on Human Rights and Technology

To: *The Australian Human Rights Commission*
From: *Responsible Technology Australia*

Responsible Technology Australia (RTA) would like to thank the Australian Human Rights Commission (AHRC) for the opportunity to contribute and input to the questions proposed in the Discussion Paper on Human Rights and Technology. We are excited with the movement on this issue as we believe it is imperative in pushing this conversation forward and ensuring appropriate and considered legislation that both protects Australian citizens whilst ensuring that we remain technologically competitive.

WHO WE ARE
RTA is an independent organisation committed to ensuring a just digital environment. We seek to ensure the safety of Australian citizens online whilst advocating for a free business ecosystem that values innovation and competition. In particular, we are concerned with the unregulated environment in which digital platforms currently exist and advocate for a considered approach to address issues of safety and threats to democracy whilst ensuring economic prosperity.

EXECUTIVE SUMMARY

We welcome the AHRC's comprehensive and macro-level approach to this issue, in considering the many direct and nuanced effects AI-informed decisionmaking might have on our society. In responding to this Discussion Paper, it is important to qualify the scope and focus of our organisation and how it fits into the broader goals of this consultation. Whilst RTA is primarily concerned with how data and technology impacts Australian democracy (in particular how the digital platforms have co-opted user data collection to fuel a pervasive 'attention economy'), AI-informed decisionmaking and the other objects of this consultation play a vital role in this mission.

As such, we applaud and agree with the AHRC's proposals to ensure that an approach that is:

Cohesive
- Through establishing a National Strategy on New and Emerging Technologies

Researched and Resourced
- Through the establishment of task forces to conduct inquiries into various impacts AI-informed decisionmaking might have on society
- Through the establishment and funding of dedicated research centres

Flexible
- Through instigating a 'regulatory sandbox' to experiment with different approaches
- Through the moratorium on facial-recognition technology until satisfactory frameworks can be put in place

Additionally, to ground and operationalise the many proposals/initiatives needed, we recommend:

- that we consider how an independent regulator could have the remit and be correctly resourced with adequate technical expertise to assess, mitigate and safeguard us from potential harms caused by new and emerging technologies including;
    - ensuring how might compliance be enforced both domestically and internationally and
    - how measured and proportional enforcement might be implemented

1.0 CONTEXT

RTA is principally concerned with how AI-informed decisionmaking has been co-opted by the large digital platforms and used to fuel an attention economy that has progressively deteriorated democracy, furthered polarisation and left society and communities vulnerable.

It is important to recognise that the profit models of digital platforms such as Facebook or YouTube are predicated on the capture and maintenance of user attention. This 'attention economy', which directly monetises the amount of time consumers spend on the platforms is optimised through the unfettered and limitless collection of personal data.

As the digital platforms have been building a comprehensive profile of their users that encapsulate their interests, vices, triggers and vulnerabilities, their algorithms have used this information to feed tailored content that is calculated to have the greatest potential of keeping users engaged. This content has been shown to lean toward the extreme and sensational, as it is more likely to be captivating[1,2].

From foreign interference in our democracy, the amplification of fringe and extremist voices that drive division, to threats to the safety of our children, the societal harms caused by this unfettered and unregulated system has permeated our society. In particular, the capacity for granular targeting of specific communities and even individuals gives rise to a completely unprecedented landscape. Whilst these harms sometimes fall outside of what is considered illegal, their negative effects on an Australian way of life are clearly evident.

Harm to Society	Example of Harms
Foreign Interference	A network of Facebook pages run out of the Balkans profited from the manipulation of Australian public sentiment. Posts were designed to provoke outrage on hot button issues such as Islam, refugees and political correctness, driving clicks to stolen articles in order to earn revenue from Facebook's ad network[3]. A number of the same accounts Twitter identified as suspected of operating out of the Russian Internet Research Agency (IRA) targeted Australian politics in response to the downing of flight MH17, attempting to cultivate an audience through memes, hashtag games and Aussie cultural references[4].
Amplification of Fringe and Extremist Voices	Datasets were collected from six public anti-vaccination Facebook pages across Australia and the US, with it appearing that although anti-vaccination networks on Facebook are large and global in scope, the comment activity sub-networks appear to be 'small world'. This suggests that social media may have a

[1] Vosoughi et al. (2018), 'The spread of true and false news online', *Science* found at https://science.sciencemag.org/content/359/6380/1146

[2] Nicas (2 Feb 2018), 'How YouTube Drives People to the Internet's Darkest Corners', *Wall Street Journal* found at https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478

[3] "Facebook trolls and scammers from Kosovo are manipulating Australian users" by Michael Workman, Stephen Hutcheon, ABC News (Mar 16, 2019)

[4] "Russian trolls targeted Australian voters on Twitter via #auspol and #MH17" by Tom Sear, Michael Jensen, The Conversation (Aug 22, 2018)

	role in spreading anti-vaccination ideas and making the movement durable on a global scale[5].
Safety of Children	A leaked Facebook document prepared by Facebook Australian executives outlines to advertisers their capability to target vulnerable teenagers as young as 14 who feel 'worthless', 'insecure' and 'defeated' by pinpointing the "moments when young people need a confidence boost" through monitoring posts, pictures, interaction and internet activity in real-time[6].

Table 1: A selection of examples of societal harms caused by an unregulated attention economy

As such, we are highly supportive of the AHRC's process to further understand AI-informed decisionmaking and its effect on society, and how we might begin to put the requisite human rights and public safety guardrails in place.

2.0 OVERARCHING RECOMMENDATIONS

In order to achieve a regulatory environment which has the flexibility and capacity to ensure a safe and prosperous online environment for all Australians, the often competing demands of innovation and protection must be balanced.

RTA supports a principled approach to regulation, where industry abides by a rigorous code of conduct but companies are free to design and distribute products and services in line with Australian values and expectations on safety. This also means that Government regulators must be adequately empowered to intervene in cases where this is breached, incentivising adherence in the first place. As such, the AHRC's overall approach - the combination of law, co-regulation and self-regulation is strongly supported by RTA.

In looking at the broad suite of proposals in the Discussion Paper, many recommendations are highly aligned with our organisation's aims to ensure a cohesive (through the development of a National Strategy), researched (through a human-rights impact assessment, and the independent task forces) and flexible (through the regulatory sandbox) approach. As such, the two recommendations we have put forward highlight how we think it might be possible to coordinate, implement and operationalise the broad range of initiatives/policies that have to occur to begin to make headway on this issue.

---

[5] "Mapping the anti-vaccination movement on Facebook" by Naomi Smith & Tim Graham, Information, Communication & Society
[6] "Facebook targets 'insecure' young people" by Darren Davidson, The Australian (May 1, 2017)

KEY RECOMMENDATION | A RESOURCED INDEPENDENT REGULATOR

Recommendation: The Government should undertake a process that explores whether an independent regulator whose role is to evidence and assess the harms of emergent online technologies should be newly created or incorporated into an existing structure/body through an expansion of powers and remits (Office of the eSafety Commissioner, ACCC, ACMA or otherwise).

There should be strong considerations to consolidate these powers into a fully independent regulator. Whilst there might be natural alignment for this to be housed within the Office of the eSafety Commissioner, due to the pervasive ways that unfettered user data collection and algorithmic amplification affect our society, it can be seen that there are significant overlapping responsibilities with (but not limited to) the ACCC, ACMA, Defence, Australian Intelligence Community, and Attorney-General's Department. Consolidating responsibility within a centralised and independent body will ensure that coordination and delivery is timely and efficient. This incorporates

The possibility of the creation of a new entity, adequately equipped, empowered and resourced (most likely through an industry levy that takes into account factors such as size and scope of impact) to deal with the current and evolving harms should be explored. Whilst there are benefits to this approach such as allowing the independent regulator to better consolidate knowledge and learnings across Government portfolios and functions, be properly equipped to liaise with the civil, academic and private sector, and house the necessary technical expertise for governance, research and enforcement. There are also many risks, such as the inefficiencies and potential loss of skill in starting up a new Government body and the lack of clarity in how this new regulator would interact with existing bodies. This needs a proper assessment of how best to enable this regulatory system.

Recommendation: The independent regulator must be resourced to conduct ongoing and proactive research and auditing of how the algorithms systems amplify content to users (also referred to as 'active monitoring'), focusing on the spread of harmful or divisive content. This should include not just hosting services, but ancillary services that aggregate and distribute content. Additionally, this should be done in coordination and conjunction with independent and academic research centres (that ideally would be set up with significant Government commitment).

This regulator must commit to begin to build out the evidence base of the impacts on Australian society that these algorithms are causing. This is vital to both inform future regulation and implementation.

Topics of immediate concern might include:

- Nature of age-appropriate content delivery to children, including violent and sexual material that is not illegal
- Investigate the nature of algorithmic delivery of content which is deemed to be fake news or harmful disinformation
- Audit the extent of algorithmic delivery on the diversity of content to any given user - to investigate filter bubbles
- Audit of the amplification of extremist or sensationalist content by these algorithms

KEY RECOMMENDATION | ENFORCEMENT

Recommendation: Investigate how the powers of an independent regulator might be expanded to incorporate adequate and proportionate enforcement.

To be effective, a regulator must be able to enforce regulation and go beyond setting transparency reporting expectations. A primary concern with many of the enforcement measures rely on voluntary industry compliance with the proposed expectations. We believe that in order for human rights and public safety to become embedded, there must be a commitment for the Government to display leadership and ensure these expectations are adhered to.

Enforcement should incentivise companies to comply whilst providing clear guidelines on how sanctions for non-compliance would be proportionate based on the size of the entity, scale and impact of their potential non-compliance and damage to society.

A wide range of tools could be employed and may include:
- Publishing public notices
- Enforce transparent public reporting
- Issuing provider warnings, reprimands and/or enforcement notices
- Serving civil fines and sanctions

Recommendation: Commit to developing a process that empowers the independent regulator to take action against entities without a legal presence in Australia.

There is an opportunity for the Australian Government to take a world leading role in developing new legal and legislative approaches to adequately deal with the global nature of this issue. It is vital that our independent regulator works with other governments from around the world to coordinate as only an international approach will ultimately be able to mitigate these harms.

What this might look like:
- Setting up multilateral working groups with similar entities internationally
- Adapting a similar concept to the EU's GDPR of having a 'nominated representative' to notionally help enforce compliance

3.0 CONCLUSION

RTA acknowledges the scale of the task ahead to design and implement the processes and institutions that will safeguard human rights in the face of emerging and novel technology with the ultimate aim of mitigating the societal harms they may/may not inadvertently cause. We look forward to working together to bring about the best outcomes for businesses, consumers and society at large.

Should the AHRC have any further questions or require ongoing comment, we would be happy to continue to engage.

Regards,

Responsible Technology Australia

For any further comment or clarification, please direct enquiries to:

████████████

Director of Policy | Responsible Technology Australia

████████████████████