Policy Submission

Submission on the Australian Code of Practice on Disinformation

To: *DIGI*
From: *Reset Australia*

Reset Australia would like to thank DIGI for the opportunity to input on the proposed Australian Code of Practice on Disinformation.

Reset Australia is an independent, non-partisan organisation committed to driving public policy advocacy, research, and civic engagement agendas to strengthen our democracy. We are the Australian affiliate of Reset, the global initiative working to counter digital threats to democracy. As the Australian partner in Reset's international network, we bring a diversity of new ideas home and provide Australian thought-leaders access to a global stage.

We look forward to engaging with DIGI, the wider tech industry, the ACMA and other relevant government regulators and departments through this consultation and beyond, as we push this conversation forward to ensure appropriate and considered legislation that protects Australian institutions, citizens and democracy.

Overview

We look forward to working with DIGI and other stakeholders, in particular the Government, to develop a workable Code that aligns with modern best practices around tackling disinformation and works in the interests of the Australian public.

It seems clear to us that this Code has been more of a rhetorical exercise, rather than a real attempt to safeguard Australian communities and democracy against disinformation. International precedent paints a bleak picture of the impact that an opt-in, buffet-style Code with no enforcement measures will have on driving the change necessary to serve Australians.

Our submission should both serve as a signal of our disappointment, but also a rallying call to the ACMA to consider more innovative, progressive and world-leading approaches that incorporate transparent monitoring, enforceable measures and proper public oversight. We implore that within their subsequent assessment of this proposed Code and report to Government, they consider the points we have raised, international experiences but most importantly, the real harms already experienced by the Australian public via disinformation.

1.0 Context

The use of services provided by the major digital platforms have become ubiquitous to the Australian way of life. With over 85% of Australians using social media 'most days',[1] the role of the digital platforms such as Facebook (Instagram, WhatsApp, Facebook), Twitter, Snapchat, TikTok and Google (YouTube and Google) play in our society has become fundamental to how we live, work and entertain ourselves.As this was happening, our understanding of the direct and indirect harms of disinformation and its root causes have also become clearer. Whilst the role of the digital platforms in the propagation of disinformation was catapulted into public consciousness through Brexit and the 2016 US Presidential election, the impacts have taken on a new dimension with the current covid-19 global pandemic. The spectrum of real world harms from disinformation now range from driving polarisation to public health risks.

How has this been allowed to occur? In order to arrive at a workable Code that will truly address the impacts of disinformation in our society, we must first recognise the system that has been constructed that has allowed for this phenomenon to arise. We also recognise that whilst this issue sits within a much wider and complex socio-political landscape and that the propagation of information (both online and offline) intended to manipulate, coerce and deceive others for economic or political gain is nothing new, we must stress that at no other point in human history has anyone had the ability to target and reach the population in the same way. Thus, this potential for harm represents a truly existential threat to our society, and thus a proportionate level of recognition, commitment and gravity must be had in its response.

The business models of the digital platforms have a single objective - to capture and maintain user attention in order to maximise advertisements served and profits generated. As such, the algorithms which dictate the content and information we consume are optimised to fulfil this objective, resulting in an economy that has commoditised attention. To feed this machine, the platforms have built a system of unfettered and limitless personal data collection, building comprehensive profiles of their users that encapsulate their interests, vices, political leanings, triggers and vulnerabilities. This data is then used to predict our engagement behaviour, constantly calculating what content has the greatest potential for keeping us engaged. This content has been shown to lean towards the extreme and sensational, as it is more likely to have higher engagement[2,3].

Whilst not intended in its inherent design, this system has had wide-ranging effects on our society. From the breakdown of public discourse due to increasing 'filter bubble' polarisation to the manipulation of communities through this online architecture by malicious actors, the myriad of issues can be collectively characterised by the breakdown of our 'online public square' - fracturing social cohesion, decreasing trust in government and halting productive civic debate, and more recently our public health. Disinformation has been a prominent way in

[1] Yellow Social Media Report (2020) Part One: Consumers. Found at:
https://2k5zke3drtv7fuwec1mzuxgv-wpengine.netdna-ssl.com/wp-content/uploads/2020/07/Yellow_Social_Media_Report_2020_Consumer.pdf
[2] Vosoughi et al. (2018), 'The spread of true and false news online', *Science* found at
https://science.sciencemag.org/content/359/6380/1146
[3] 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

which this has manifested, with the scale, scope and mechanisms of how it wreaks its harms still evolving.

Our Recommendations

Whilst we will be drawing out specific recommendations and opinions across the proposed Code's Objectives and Administration, our overarching concerns centre around the inherent limitations of an opt-in self regulatory Code. We are skeptical that this model will be either impactful or effective in actually tackling disinformation.

We are especially discouraged by the stark similarities between this Code and the EU's Code for Disinformation, and the lack of progress that has been made to improve upon European efforts. As mentioned in their Assessment of the EU Code of Practice on Disinformation[4]:

> *At present, it remains difficult to precisely assess the timeliness, comprehensiveness and impact of the platforms' actions, as the Commission and public authorities are still very much reliant on the willingness of platforms to share information and data. The lack of access to data … (along with) the absence of meaningful KPIs to assess the effectiveness of platform's policies to counter the phenomenon, is a fundamental shortcoming of the current Code.*

Although we broadly support the Objectives of the proposed Australian Code as they provide a framework for concerted action on disinformation, the lack of effort to incorporate learnings from the EU and other jurisdictions foreshadow a similarly ineffective implementation here, as well as adding to our misgivings around the intent and commitment potential Signatories of this proposed Code actually have around tackling disinformation.

Definition of Disinformation

This Code's definition of disinformation excludes 'misleading advertising… or clearly identified partisan news and commentary'. We strongly believe that this Code must align with best practices on truth in political advertising and strive to add to this discussion.

Additionally, the definition of harm as 'imminent or serious threat', will limit the scope for action of this Code. Whilst misinformation and disinformation practices do have imminent consequences as evidenced by the Covid-19 crisis, the true impacts on democracy, governance, public trust and social discourse can also be characterised as a slow degradation. Hence, this definition of harm as 'imminent' limits the ability of this Code not just to deal with this harms, but to enable a research and regulatory environment that is adaptable enough to deep-dive into these impacts.

---

[4] European Commission (2020), 'Staff Working Document: Assessment of the Code of Practice on Disinformation - Achievements and areas for further improvement'. Found at:
https://ec.europa.eu/digital-single-market/en/news/assessment-code-practice-disinformation-achievements-and-areas-further-improvement

> **Recommendation:** Change the definition of harm to recognise the holistic and long-term impacts of disinformation and work towards developing a framework to truly understand and assess these harms

Code Objectives

Overall, we support the six objectives that are included within the proposed Code. The commitments laid out under these Objectives lay out clear and agreeable goals that, if actioned, will begin to address the disinformation phenomenon. The specific wording of the Objectives under the Australian Code, like the Pillars within the EU Code are largely unobjectionable.

Our apprehension primarily arises from the limitations experienced by the Europeans in effectively implementing their Code, and what that implies for the Australian experience.

Objective 2 - Disrupt advertising and monetisation incentives for Disinformation

We welcome efforts this proposed Code makes to address issues within the fundamental profit models of these digital platforms that have allowed for the propagation of disinformation. This Objective represents a key opportunity to address the underlying financial drivers that are used to propagate disinformation.

As highlighted within the EU's Second Assessment of their Code of Practice on Disinformation, inconsistent implementation of measures intended to address placement of advertisements on platforms' own services limited progress against this commitment. Additional challenges were seen regarding implementation of measures intended to limit ad placements on third-party websites that spread disinformation. Furthermore, the Assessment goes on to state that 'the Code does not have a high enough public profile to put sufficient pressure for change on the platforms in this area'. These limitations were largely put to ineffective participation and collaboration by relevant stakeholders, including the advertising sector, fact-checking organisations and the platforms themselves[5].

The financial drivers which propagate disinformation represent the key opportunity for initial action, and this Objective is a valuable first step in the recognition of these responsible economic incentives. However, for this Code to actually be a code of 'practice' more guidance and work must be done to assist parties to actually operationalise these commitments.

---

[5] European Commission (2020), 'Staff Working Document: Assessment of the Code of Practice on Disinformation - Achievements and areas for further improvement'. Found at: https://ec.europa.eu/digital-single-market/en/news/assessment-code-practice-disinformation-achievements-and-areas-further-improvement

> **Recommendation:** Work towards developing better defined guidance, practices to collaboration pathways that will effectively disrupt the economic drivers of disinformation, moving beyond the broad commitments listed under the current Objective. This should include:
> - Developing a common structure for risk assessment and escalation framework for ad accounts that propogate disinformation
> - Developing an application-approval system for actors intending on using advertising based on agreed-upon trustworthiness indicators
> - Defining concrete ways in which transparency can be embedded in on-platform advertisements to users, as well as wider transparency measures to the public and relevant stakeholders to ensure accountability.
> - Defining pathways for greater collaboration with other relevant stakeholders, in particular the advertising sector.

Objective 5 - Strengthen public understanding of Disinformation through support of strategic research

Although we are broadly supportive of the outcomes and measures listed under this Objective, we are skeptical of how these commitments will actually play out in reality. Whilst it is presumptive to assume how this Code will actually play out in the Australian context, we are already disappointed in the proposed wording of this Objective as it belies what we perceive to be disingenuous intent to actually support a research agenda in this space.

This Objective largely lines up with Piller E of the EU's Code of Practice on Disinformation around empowering the research community. However compared with the European code, there has been an omission of a specific commitment towards encouraging research into disinformation and political advertising that is found within the European Code. Additionally, in the second assessment of the EU's Code found that the goals under Pillar 5 had largely not been achieved, with a 'shared opinion amongst European researchers that the provision of data and search tools required to detect and analyse disinformation cases is still episodic and arbitrary, and does not respond to the full range of research needs'. Specifically, the discretionary approach the platforms take of entering into bilateral relationships with specific members of the academic and fact-checking community flies against the open and non-discriminatory approach needed for the levels of research, analysis and accountability required. Even for actors who do engage with the platforms for research purposes, this tight-grip on data leads to a power imbalance which calls into question the integrity of outputted research.

As the nature and impacts of disinformation are still evolving, our shared understanding of this issue through proper research is more important than ever which is why international experiences such as with Pillar 5 of the EU Code prove especially disheartening. This Code purposefully omits a key commitment found within the EU Code, which further weakens potential action, and if taken optimistically, the commitments listed under Objective 5 of the proposed Australian Code have failed to utilise any learnings from 2 years of implementation in other jurisdictions to come up with any new and/or innovative ways to strengthen research.

> **Recommendation:** Each Signatory must commit to developing data sharing arrangements that empower academic researchers, civil sector actors, think tanks and public regulators to undertake the requisite research on disinformation. These arrangements must preserve user privacy but also make good faith attempts to increase transparency on data that is vital for our understanding of disinformation (e.g. demographic data of user engagement, content engagement).
>
> An example of a proposal for such an agreement can be found in a policy memo we developed called the [Data Access Mandate for a Better COVID-19 Response in Australia](). Whilst this memo focuses specifically on COVID-19 disinformation, this transparent data access proposal can and should be extended to other areas of disinformation research that impact our community.

Code Administration

Oversight

In order for this proposed Code to have the legitimacy required to ensure that progress is aligned with its commitments, implementation must be administered by a party that is both objective and has public interest prioritised. As an 'industry association representing the digital industry in Australia'[6], DIGI does not fulfill either of these requirements and thus must either be a Co-Administrator or defer duties to a more impartial body.

An independent third-party organisation and/or commission of relevant stakeholders must be established as Administrator of the Code, and must come from sectors that represent key stakeholders in the rollout of this Code such as civil sector, academia, media and/or the general public.

> **Recommendation:** The third-party organisation chosen to be the Administrator of this Code must be independent, objective and prioritise public interest

Regarding the sub-committee established to review and monitor the actions and commitments of Signatories, this must not be established at the sole discretion of DIGI for similar reasons mentioned above, and be made up of a broad and diverse range of representatives. Additionally, members of this sub-committee must transparently declare any conflicts of interest and ensure best practice governance standards are followed. We recognise that DIGI will play a pivotal role in convening and incorporating views from the industry, and thus see the organisation playing a role here, however in conjunction with other representatives from sectors such as civil, academia, government, media and others.

---

[6] https://digi.org.au/about/

> **Recommendation:** Ensure that the sub-committee established by the Code Administrator to monitor and review the actions and committees of Signatories be made up of diverse range of stakeholders comprised key impacted representatives and respected experts

As the impacts of disinformation are borne by the Australian society, there must be a recognition of the fundamental role government bodies and regulators will play in combating this issue. Clear mechanisms that will embed cooperation between Signatories and these respective bodies must be detailed within this Code. These mechanisms must also illustrate how regulators such as the ACMA can input into the review and compliance of commitments, as well as any future development and/or iterations of the Code.

> **Recommendation:** The ACMA must play a leading oversight role in the Code's implementation and operationalisation. Public oversight and regulation must be evidenced in the Code's governance structure, reporting requirements and implementation plan. This public oversight must be included into the Code, including establishing clear mechanisms of collaboration, governance and review between government regulators (the ACMA), Code Signatories and the Administrator of the Code.

Monitoring, Reporting and Review
We welcome the Code's commitment through Annual Reporting to develop and implement an agreed format to report on progress against the Commitments, however insufficient detail was found within the proposed Code on planned monitoring and reporting efforts. A commitment to actually understanding how this Code will impact the information disorder landscape in Australia starts with resourced and rigorous monitoring and reporting, otherwise this exercise will be redundant and bureaucratic.

A key finding within the 2nd EU Assessment on their Code of Practice on Disinformation revealed 'substantial deficiencies, owing in part to the lack of a common reporting structure and a consistent understanding of monitoring and evaluation needs'. Their Assessment states that some of these barriers related to the multi-jurisdictional market of the EU, hence we assume that developing a cohesive monitoring and reporting system for the Australian market would be less complex.

Whilst we respect the need for an individualised approach for each platform, there needs to be consensus on how the Objectives laid out in the code translate to a shared understanding of what progress looks like in this issue. Currently, the proposed Code, has made no real attempts to actually drive progress on this issue, failing to provide answers to questions such as:

- How they will actually assess disinformation, its harms/impacts?
- How they will design their safeguards against disinformation, what their measure of proportionality is?
- How does their list of proposed measures actually apply to different types of disinformation, their harm/intended harm and/or perpetrating actor?

Additionally, the ACMA Guidance specifies that key performance indicators and a reporting scheme should be developed and has offered to collaborate with the platforms in order to achieve this.

> **Recommendation:** Work with the ACMA to develop and define a common reporting structure and shared KPIs that will able to adequately monitor the implementation of Signatories commitments and measures

We echo the EU Assessment's findings on the need for two distinct classes of KPIs when developing out the proposed Code's monitoring system:
- Service-level Indicators - in order to measure and assess the results of specific policies and initiatives implemented by Signatories, whilst accounting for differences between platform services
- Structural Indicators - to measure the prevalence and impact of disinformation at a societal level, in order to assess the impact of the Code on disinformation within Australia

In particular, to make significant progress towards understanding the structural indicators (arguably the most important aspect), there should be a consideration for greater efforts to address our lack of understanding.

This would take the form of expanded responsibilities of an independent regulator, which would work to build an evidence base on how algorithms prioritise and distribute certain content and the impact of this on society in order to inform future regulation. Algorithmic audits of these platforms would need to involve mandatory collaboration with the relevant companies. This should focus on (but not be limited to) the following:

- Investigate the nature of algorithmic delivery of content which is deemed to be fake news, propaganda or disinformation
- Audit the extent of algorithmic delivery on the diversity of content to any given user to investigate the impact of ideological filter bubbles
- Audit of the amplification of polarising and extremist political content by these algorithms
- Understand demographic nuances for all of the above, particularly as it relates to the specific targeting of diaspora communities in Australia.

An example method to explore modelling algorithmic auditing is available at Algo Transparency, which provides a snapshot of the videos recommended on Youtube.

> **Recommendation:** Explore how ongoing and proactive auditing of the content that algorithms amplify to users, focusing on the spread of disinformation and its impacts might be incorporated into the future direction

Guidance on Platform-specific Measures

We strongly support the position stated within the ACMA's position paper intended to guide the Code's development on a harms-based approach when deciding what proportionate measures to apply to address disinformation. We also recognise the diversity in the range of actions and measures that might be employed to address the range of harms and that an approach that is over-prescriptive would be counterproductive.

However, we strongly encourage the proposed Code to formalise a harms-based framework to potential measures, rather than just providing a list of potential actions that 'might' be employed. We emphasise the point raised in the ACMA's Guidance, that providing an industry-wide guidance on assessing potential risks would promote consistency across the industry and future-proof the Code by allowing platforms to adjust their measures in response to new developments. We would also like to reiterate that research into detecting, assessing and mitigating disinformation is still nascent, and that efforts to build a common understanding of the harms of disinformation would provide the bedrock for other sectors (such as academia and civil society) to build upon, and a much needed signal to the wider Australian public on the industry's commitment to addressing these harms.

> **Recommendation:** Develop and incorporate an industry-wide framework for assessing the risk for harm of disinformation, and formalise guidance on mapping specific measures to these risks

Conclusion

We strongly believe that a self-regulatory code will be insufficient in addressing the harms that arise through disinformation. Whilst we support the Objectives and align on what the authors of the Code perhaps hope they would achieve, the likely reality based on numerous international examples as well as other efforts the platforms have made within our country to limit public interest tech regulation paints a different picture.

As an opt-in, take-as-you-see-fit policy, with no real negative effects for ignoring the Code of Practice altogether, we believe we are at an impasse until the Government steps in.

For any further clarification or comment, please contact:

Matt Nguyen
Policy Lead
matt@reset.tech.au