Initial analysis of Misinformation / disinformation concerns

Note: These concerns are indicative from initial analysis, and not exhaustive.

The definition of misinformation and disinformation is broad

The definition covers content that is "reasonably verifiable as false, misleading or deceptive" and "reasonably likely to cause or contribute to serious harm". Disinformation is when "there are grounds to suspect" that the person "intends [to] deceive another person" or "the dissemination involves inauthentic behaviour" (13(1-2)).

"Serious harm" is given a lower and broader standard than other legislation and includes "vilification" of a group. Vilification is not defined, though concerningly, the Explanatory Memorandum cites the *Anti-Discrimination Act 1992* (NT) as an example of a vilification prohibition. This prohibition includes the extremely low bar of "reasonably likely to offend". The Memorandum states that the inclusion of vilification as a harm in the Bill "aligns broadly with the approach" taken in the NT, Act and other State based anti-discrimination acts.

This creates an extremely low standard of "serious harm" that includes "reasonably likely to offend".

The exemptions for speech are limited and subjective

The exemption for religious speech is limited to "reasonable dissemination". The standard of "reasonable" has been recognised to be highly subjective.[1] The explanatory memorandum says that the "reasonable" test attaches to the dissemination and not the substance of the content. *However*, the memorandum goes on to say "if content is in the first place so unreasonable, it will also be unreasonable to share it any further."

This leaves the exemption entirely dependent on what a judge, or content provider, determines is "reasonable".

Private companies are expected to interpret and apply this law

The breadth of definition and the subjective nature of the exemptions are problematic when interpreted by a judge. They are *extremely dangerous* when left in the hands of private companies famed for their opaque processes and rapacious profit-making. These companies become the judge, jury and executioner of content while claiming the mandate of Government.

These companies are incentivised to take as broad an interpretation as possible to avoid legal repercussions (including ACMA enforcing new standards). The nuance of what is or is not "harmful" speech, or "reasonable" religious expression will be lost behind caution, profit, and ideology.

---

[1] E.g., comments by French J and Gleeson CJ in 'Bropho':
Bropho v Human Rights & Equal Opportunity Commission [2004] FCAFC 16 (6 February 2004) [76]
Bropho v Human Rights & Equal Opportunity Commission HCA Transcript 9 (4 February 2005)

Automatic processes will automatically censor religious speech

Due to scale, the only option that digital providers have is automatic AI-based and rules-based censoring. To avoid false negatives that allow "misinformation" to remain, the systems will be biased towards false positives and will falsely detect content as "misinformation".

There is no transparent means of appeal or recourse

Users of digital platforms already report great difficulties in appealing the removal of their content. Adding in the legal risk of non-compliance, and the ability to blame Government policy for over-zealous censorship, will make the process even harder. Many platforms also have mechanisms for "shadow-banning" where a user is not notified of censorship, but their content is prevented from being seen by others.

Additionally, while digital providers are required to provide some form of reporting, there does not appear to be a mechanism for knowing what content has been censored and for what reason.

At minimum, this legislation should require full transparent reporting of the content that has been censored, and an independent enforceable appeals process, including

- immediate notification to users when their content has been removed or restricted
- regular reporting by providers of all content that has been restricted – made available to Government and researchers
- independent and transparent appeals process, including the ability to exercise the exemptions in the Act to demonstrate that content is not misinformation.

Our Concerns

It is reasonable to expect that, out of an abundance of caution, a digital provider will take the most expansive interpretation of this law. It is easy to imagine a scenario where legitimate debate about religion or ethics can be considered by a provider (either via an algorithm or a busy "fact checker") to be "reasonably likely to offend", or otherwise "harmful", and not a "reasonable" expression of faith. Examples include:

- Debates between faiths where one faith member declares that another faith is wrong and "going to hell"
- Debates on whether a faith's ethical standards on gender or sexuality are oppressive
- Statements about whether or not a man can transition to become a woman
- Discussions about whether sexual expression is right outside of heterosexual marriage

It is also easy to foresee that activist campaigns against these corporations (that are wholly dependent on advertising for revenue), or internal ideological pressures, would tip the balance of censorship to one side or the other of the ideological spectrum.

In that inevitable scenario, there is no mechanism for independent review and appeal, and no data available for detecting systematic failures.

It is difficult to see how tinkering with the definitions or exemptions will correct the problem. The fundamental issue is that profit-focused, risk-adverse and politically manipulatable private corporations are given the mandate to monitor speech with significant financial penalties for not being strict enough, no penalties for being too strict, and no mechanisms to defend freedom of speech and enforce strict adherence to definitions and exemptions.

Mike Southon · Executive Director, Freedom for Faith · mike@freedomforfaith.org.au · 0412 651 660