



Submission on the Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023

Thank you for the opportunity to make a submission on this Exposure Draft Bill.

The Institute for Civil Society (ICS) is a social policy think tank. ICS seeks to:

- 1. Promote recognition and respect for the institutions of civil society that exist between individuals and the government.*
- 2. Uphold traditional rights and liberties, including the freedoms of expression, association, conscience and belief.*
- 3. Promote a sensible and civil discussion about how to balance competing rights and freedoms in Australian society.*

Section 1 - Philosophical/ethical basis for freedom of expression

In this section, we outline the key philosophical concerns raised by the Bill through the lens of 'free speech' (or more broadly, freedom of expression). Freedom of expression is a human right which the Australian government has recognised as a right possessed by all people. Australia is a signatory to the International Covenant on Civil and Political Rights, art 19 of which provides that:

- Everyone shall have the right to hold opinions without interference.
- Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds,

regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.

Freedom of expression is widely regarded as crucial for (1) human dignity (2) the ascertainment of truth and (3) the effective functioning of a democracy. This section of our submission will explain these justifications and explain how each one is unjustifiably compromised by this Bill. The concerns raised will be further contextualised in reference to specific provisions of the Bill in Section 2.

Freedom of expression is essential to human dignity.

Free speech is valued on the basis that it is essential to human dignity. This justification is not outcomes based. Free speech is valued simply because speaking and expressing is what makes us human. Accordingly, restricting human beings in how they express themselves deprives them of their full personhood. The justification goes so far as to contemplate a right to be wrong. While freedom of expression is a universal human right, it does not follow that all persons have equal access to the channels of communication. In their book *Tell Our Story: Multiplying voices in the news media*, Reid and McKinley explain that the media can (and do) marginalise or increase the volume, and therefore reach, of different voices.

The advent of social media has given people the opportunity to be citizen publishers on a massive and unprecedented scale and to that extent has democratised communication. Indeed, for better or worse, social media has become an indispensable means of expression and communication for a large percentage of the population, as well as their primary source of news. However, depending on the biases of the social media platform and its degree of commitment to free speech, some segments of society are more likely to find their speech curtailed than others. This is the case even without the enactment of the Bill.¹ For example, a peer reviewed study conducted at the University of Michigan concluded that while 'social media sites use content moderation to attempt to cultivate safe spaces with accurate

¹ Twitter (now renamed X) was purchased by Elon Musk with an accompanying commitment to allowing greater free speech on the platform. The ABC has recently closed most of its accounts on X, citing toxic interactions as the reason. Musk responded to the national broadcaster's decision by accusing it of preferring 'censorship-friendly social media'. Irrespective of where the truth of the matter lies, the perception is that some social media platforms are more free speech friendly than others.

information for their users' this moderation 'may not be applied equally for all types of users, and may lead to disproportionate censorship related to people's genders, races, or political orientations.'²

Upholding people's dignities through their right to freely express themselves will also help prevent group think and inhibit the growth of extremist views, particularly in the long run. If group think prevails, people lose their tolerance of people with differing opinions to them, a phenomenon already prevalent in our society. This notion of 'othering' can clearly be seen in instances of 'cancel culture' for example.

By requiring the creation of codes which are designed to identify 'harmful' 'misinformation' and 'disinformation' and imposing huge fines on social media companies if they do not remove such content, the Bill will frustrate self-expression on platforms that have become a major means of communication and interaction for many people, thereby undermining their human dignity. Before enacting the proposed legislation, the government needs to consider whether the consequences of such an approach to 'harmful' 'misinformation' and 'disinformation' outweigh the benefits, in the short and long term. We argue that they do not.

Freedom of speech is an integral part of the ascertainment of truth.

This argument for free speech is more utilitarian. It postulates that if all persons are free to make their claims and express their views, the truth will ultimately win out in the marketplace of ideas.³ Yet instead of permitting the marketplace to expose false claims and attack weak arguments, the Bill would vest this task in social media platforms, which, as noted above, may have their own biases and which, in any event, are simply not equipped to make decisions on every contested claim. Moreover, the removal of information deprives the public of the opportunity of undertaking the disciplined and rigorous process involved in analysing information and of determining for themselves its merit. Removal of misinformation or disinformation prevents the public from accessing the information, leaving them with nothing to analyse. The silencing of particular voices online is particularly concerning

² Oliver L. Haimson et al., "Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas," *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW2 (October 2021): 1, doi:10.1145/3479610.

³ Cite Mill etc

in view of the fact that what has been deemed 'harmful' 'mis/disinformation' might be found at a later date to be true. A case in point is the Hunter Biden computer saga. When the story first broke, over 50 former US intelligence officers dismissed it as Russian disinformation designed to influence American voters in the forthcoming election; it has since been proven true. Had the story been removed or suppressed, the American public would have been deprived of access to material in which it had a legitimate interest.

Free speech is crucial for active participation in a democracy.

A true democracy is 'rule by the people', which is implemented through a system of government in which citizens elect members of parliament who act as their representatives. The right to directly choose our members of parliament is enshrined in the Constitution. Citizens in a liberal democracy can exercise their constitutional right to directly choose their members of parliament and to change the constitution via a referendum in an informed manner only if they have access to political and government information. This notion is at the heart of the implied freedom of political communication, which was first recognised by the High Court in 1992 in *Nationwide News Pty Ltd v Wills* (1992) 177 CLR 1 and *Australian Capital Television v Commonwealth* (1992) 177 CLR 106. This justification for freedom of expression only applies to speech that has relevance to people's role as voters so it is a more limited justification, but one that has received the most attention from the High Court.

The Albanese Government's initial decision to promote only the 'yes' side of The Voice campaign and abandon the usual process of circulating arguments for both the 'yes' and 'no' sides (which in fairness they subsequently abandoned) was an affront to the reasoning underlying the implied freedom. The Yes Case proponents' behaviour during the campaign so far has been to obfuscate and to denigrate those who have sought details on how The Voice would work without any concession that this is information to which the Australian public is entitled. If social media companies and ACMA adopt a similar attitude under the Bill it does not augur well for political discourse. Removing posts before they can be critically analysed and responded to (or ordering their take down shortly thereafter) unjustifiably silences citizens' voices.

1.4 Conclusion

Freedom of expression is an essential human right, and while we recognise that it cannot be absolute, we maintain that any restrictions on it should be very compelling and in the vast majority of cases any punitive consequences should be imposed after the fact, not pre-emptively. The policy behind the Bill is that “Misinformation and disinformation pose a threat to the safety and wellbeing of Australians, as well as our democracy, society and economy.” Our stance is that the suppression of information is a greater danger to our democracy, society, and economy. To repose the right to determine what is misinformation and disinformation in faceless, unelected people and bodies is an affront to democracy. It infantilises the Australian people and betrays a lack of trust in their ability to sift the wheat from the chaff. It impinges on people’s democratic responsibility to receive and engage critically with information and with opinions that might differ to their own.

Section 2 - Banning of misinformation and disinformation (via the agency of private sector platform providers) is detrimental.

In section 1, we argued that the Bill is an unjustified affront to freedom of expression and will have an adverse impact on each one of its three main justifications. In this section we evaluate the Bill at a more detailed level. We argue that the definitions of misinformation, disinformation and harm, are too broad and raise a number of serious concerns about its impact on the freedom of Australians to make claims about, and express views on, many issues which affect them.

2.1 Definitions of misinformation and disinformation

(a) Are opinions and statements of feelings information which might be misinformation?

While the focus of the public discourse about the Bill has centred on the ‘mis’ and ‘dis’ components of misinformation and disinformation, we first consider whether the ‘information’ component has any inherent limits. For example, is an opinion information? Are expressions of feelings information? According to the Stanford Encyclopedia of Philosophy, the word information is colloquially used to ‘denote any

amount of data, code or text that is stored, sent, received or manipulated in any medium.’ Any statement posted on social media could feasibly fit into this definition. If so, an expression of opinion or the expression of emotion can count as information. If this is the case, can the expression of one’s opinions or emotions be labelled as ‘misinformation’?

There is the potential for ‘information’ to be construed in such a way that it is not strictly confined to expressions of fact. For example, the statement ‘man-made climate change is a hoax’ is presented as a statement of fact. It purports to be information, and is arguably misinformation. By contrast, the statement ‘in my opinion, man-made climate change is a hoax’ is an expression of opinion. Might it nevertheless be regarded as ‘information’ for the purposes of the Bill, since the ‘sting’ of the claim is the same in both cases? Yet it is not strictly misinformation for a person to claim that they do not believe in something. Indeed, unless the person is lying, the statement is factually accurate. If the Bill is to be restricted to statements of fact, it should say so. Otherwise, it will likely capture opinions and emotions.

(b) The definitions of misinformation and disinformation are too broad.

Misinformation is defined in cl 7(1) as:

- a) content disseminated using a digital service that is false, misleading or deceptive;
- b) the content is provided on the digital service to one or more end-users in Australia;
- c) the content is not excluded for misinformation purposes; and
- d) the provision of the content on the digital service is reasonably likely to cause or contribute to serious harm.

Disinformation is simply misinformation where the disseminating, or causing the dissemination of, the content is done with the intention that the content deceive another person.

We raise concerns with the three descriptors - ‘false’, ‘misleading’ and ‘deceptive’ - in the definition of misinformation, only one of which need be present in order for the provision to be satisfied.

- *False*

False means 'not true' and thus should only be applicable to statements that are objectively verifiable (able to be demonstrably proven to be either true or untrue). This harks back to our discussion of 'information'.

This presents an issue in our post-modern society where we have embraced the concept of 'your truth'/ 'my truth'. The prevalence of this thinking has moved us away, in some respects, from objective reality. At one level it is contradictory to promote this way of subjective thinking and at the same time to moderate mis/disinformation in such a 'black and white' way. If reality is largely subjective, then there is less basis on which to determine falsity. Of course, this is an overstatement, but it does apply here to a degree. If we want to allow Australians to identify themselves in any way they choose, for example, we cannot also dismiss their statements – or the contrary statements of others - as mis/disinformation.

- *Misleading*

The composite phrase 'misleading and deceptive' is derived from consumer law. 'Misleading' refers to information that gives the wrong idea or impression. It might catch information that is true but misleading due to its context. Its determination requires an assessment of the effect of the information on the audience. Is the social media platform required to ascertain the likely readership of the post and what else they knew when they read the post so as to ascertain its likely effect? Or, if the post is generally accessible to the general public, is the social media company required to consider whether it would mislead the 'ordinary person'? Once the definition moves away from statements that are objectively true or false, the assessment becomes more complex and subjective.

For example, take the statement 'not all mothers are women'. This may be treated as true based on certain assumptions as to the meaning of "women" and mothers" but it might be false or misleading based on other assumptions. Is this a misleading statement?

- a) “ACMA found that 82% of Australians report having seen Covid-19 misinformation over the past 18 months, warning that “falsehoods and conspiracies” online had undermined Australia’s public health response.”⁴

It is no longer clear whether some of the information that Australians reported as misinformation about Covid-19 is indeed misinformation. This statistic is also potentially misleading because it is merely referring to reports of misinformation, not whether people were actually being misled.

- b) Members of the ‘Yes’ side of The Voice campaign criticised the ‘No’ side for quoting members of the ‘Yes’ side in their official pamphlet. They claimed that their quoted statements were taken out of context⁵

- c) Whether information is misinformation may depend on timing – it may later become inaccurate because of evolving events or new knowledge. News like that spread soon after Hurricane Harvey⁶ may be true at the time, but decisions in response to a crisis are consistently evolving. An analysis of the information spread on Twitter post Hurricane Harvey found that when statements were posted too soon regarding different responses, it was hard to get rid of what was spread when that news became ‘fake’ at a later time, even hours after a decision had initially been made. Are social media companies obliged to consider the time of posting or the time of reading or both in assessing misinformation. Is stale news misinformation?

- *Deceptive*

This means to give an appearance or impression that is different from the true one. This could catch information that is presented in a deceptive manner but which is true. An example might be when other facts are concealed which would put a different gloss on the information.

⁴ “Digital Code of Conduct Fails to Stop All Harms of Misinformation, ACMA Warns | Australian Media | The Guardian,” accessed August 18, 2023, <https://www.theguardian.com/media/2022/mar/21/digital-code-of-conduct-fails-to-stop-all-harms-of-misinformation-acma-warns>.

⁵ We understand that the official No case is not a purely online publication.

⁶ So-Min Cheong and Matthew Babcock, “Attention to Misleading and Contentious Tweets in the Case of Hurricane Harvey,” *Natural Hazards* 105, no. 3 (February 2021): 2883–2906, doi:10.1007/s11069-020-04430-w.

2.2 The process for identifying misinformation and disinformation

It is unclear how misinformation and disinformation will be identified and how the two concepts will be distinguished from each other. Presumably the process will be to some extent depersonalised. We envisage that an algorithm will be employed by the social media company to flag certain words and phrases that relate to misinformation or disinformation. It is unclear whether a human fact checker will then remove the post or whether any human fact checking will only be done after the removal (if at all). This raises a number of concerns:

First, who will do the fact checking? What are their qualifications? How can they have sufficient expertise to, for example, censor a qualified doctor's comments on the efficacy of a particular medical treatment?

Second, will their reasoning be recorded and made accessible to the public (or at least to the person who posted the content)? Will there be a procedure for informing a person as to why their content has been removed?

Third, in view of the fact that the different digital platforms will be subject to different codes of practice with different standards, is there a danger that the codes will produce different conclusions, such that certain information is removed from one platform but not from another? If so, might this inconsistency trigger ACMA to impose a standard?

Fourthly, is the censorship of information entirely pro-active, done as a result of algorithmic searching, or might it also occur in response to a complaint? If it can also operate responsively, what is to stop activists from making persistent complaints about posts they dislike? They could collectively choose to bombard a digital platform with complaints. With enough people and determination, a class and or group of people could assert themselves as the arbiters of truth on a particular topic.

Fifthly, if algorithms are used, will the algorithm favour certain groups? A study conducted at the University of Michigan, *Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social*

Commented [SR1]: Needs a sub-heading.

Commented [SR2]: I have moved this to where we discuss 'false'.

Media Users: Marginalization and Moderation Gray Areas, noted that 'stories have emerged about certain groups of social media users who perceive that their content and accounts are removed more often than others.'⁷ This motivated them to search for evidence, which they found, that content moderation similar to that proposed in this Bill, disadvantaged different groups.

2.3 Review of Removals and Recourse/Reinstatement in the Case of Incorrect Removals

What happens if information is removed on the basis that it is false and the information subsequently turns out to be true? The Bill contemplates that serious harm might be caused by the presence of mis/ disinformation, but it might also be caused by reason of the suppression of that information.

During the 2020 US election campaign, allegations were made that Hunter Biden had left incriminating evidence on an abandoned computer. The allegations were dismissed as 'Russian disinformation and conspiracy theory'⁸ designed to have a detrimental effect on Joe Biden's run for the presidency. However, the dismissal was debunked, with news outlets like the New York Times admitting, only relatively recently, that the incriminating evidence was actually there. Thus, the claims of mis/disinformation were in fact mis/disinformation themselves! Now that the truth has been discovered, it is having an even larger effect on Biden's credibility because the information was suppressed for so long, those fighting for the truth being gaslit and marginalised. This only increases the polarisation existent in US politics by validating those who previously held the minority view.

Recently, the BBC published a claim to the effect that Nigel Farage was de-banked because of his financial situation. This information was provided to a BBC journalist by the bank's CEO. This has been proven untrue; Nigel Farage was actually 'debanked' for his political views, namely, his support of Brexit and Donald Trump.

⁷ Oliver L. Haimson et al., "Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas," *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW2 (October 2021): 1–35, doi:10.1145/3479610.

⁸ C. Mitchell Shaw, "Hunters Laptop Now Recognized as Genuine.," *The New American* 38, no. 8 (April 2022): 21–25.

Farage was able to produce a 40-page internal bank document to prove it. Had this internal document not come to light, the false claim made by the BBC would not have been exposed and Farage's debanking would have continued. In view of the fact that Farage claimed that other UK banks had refused his custom, serious harm would have been caused to a UK citizen had the status quo prevailed and all because a UK citizen expressed legal political views.

Will suppressed information that turns out to be true be reinstated? It appears that Codes and standards do not have to provide review mechanisms, reinstatement or compensation for wrongful removal of content but they should. And if, for example, Facebook is fined for not complying with the code by not removing from its platform 'misinformation' that is subsequently found to be true, will the fine be returned? And if, for example, Facebook *had* removed views that have turned out not to be false, misleading or deceptive, the Bill will have unjustifiably stifled the freedom of expression of those who posted them. The Bill makes no provision for compensation for those affected by wrongful removal of their content but it should. But it means that a violation of human right goes unremedied.

- What if there is an objective answer to an assertion but the answer is unknown?

"The covid virus began at a Wuhan market."

"The covid virus escaped from a Wuhan lab."

The concept of falsity cannot sensibly be applied to opinions and prognostications. However, the danger is that they will be applied to these things. Would any of the following controversial statements be suppressed as false?

- a) "We will get a public holiday if the Matildas win the World Cup."

"We will not get a public holiday if the Matilda's win the World Cup."

- b) "A foetus is a human being."

"A foetus is not a human being."

- c) "Wind turbines and solar panels do more harm to the environment than good."

“Solar and wind energy are worthwhile investments to help the environment.”

There will undoubtedly be differences of opinion as to whether the above statements are misinformation. If this is the case, they should be left on the platform for the public to express their own opinions on the matters at hand.

2.4 Definition of Harm

The proposed powers will only apply to misinformation and disinformation that is reasonably likely to cause or contribute to serious harm.

Harm means any of the following (and the examples below are given in the fact sheet):

- (a) hatred against a group in Australian society on the basis of ethnicity, nationality, race, gender, sexual orientation, age, religion or physical or mental disability;

Example: Misinformation about a group of Australians inciting other persons to commit hate crimes against that group.

How can an algorithm or platform representative decide whether misinformation will incite violence or not? Surely the responsibility belongs to the beholder who decides to incite violence. Why are existing vilification laws not a sufficient deterrent and punishment?

- (b) disruption of public order or society in Australia;

Example: Misinformation that encouraged or caused people to vandalise critical communications infrastructure.

Again, we can query whether the post and the vandalism are correlated or the post caused the vandalism. What about a call out on social media to attend a protest? If the digital platforms regard the basis for the call to protest as misleading, will the protest itself be regarded as likely to cause or contribute to serious harm? Would there need to be credible information that it would be violent? Damage property? For example, the large protests in Melbourne against mandatory vaccines and the emergency powers bill in late 2021 were no doubt disruptive to shops in Bourke St (because no one could access them in the thick crowds) and disruptive to traffic. It also had a lot of police attention but there was no violence.

(c) harm to the integrity of Australian democratic processes or of Commonwealth, State, Territory or local government institutions;

Example: Misinformation undermining the impartiality of an Australian electoral management body ahead of an election or a referendum.

The recent Craig Kelly judgment is an example of when someone was accused of this. The AEC pursued him for what, in practice, was the font size of his authorisation (too small), having photographed it from a distance in the dark and not told him where the offending posters were. It's a very damning judgment as it was clearly the AEC that had not acted with integrity. Apparently, posters and corflutes of other parties did not always comply but they were not pursued.⁹

45 Quite how Mr Kelly could have dealt with the demand to remove or amend the two signs without being told where they were in a large electorate was not explained in any correspondence by or on behalf of the Commission or in argument at the trial. As I said in the course of argument, given the seriousness of Mr Kelly's alleged contraventions which the Commission was asserting, it is difficult to understand why it did not tell him fairly and precisely where the allegedly infringing signs were. Of course, it could also have added that he had also to ensure that the rest of his posters all complied with the law. The Commission would be aware of the realities and the various campaign tasks performed by a candidate, his or her staff and volunteers, when involved in any candidate's election campaign, including the widespread dissemination and use of corflutes and other signs over the geographic spread of an electorate. Not informing a candidate or party of the location of allegedly contravening conduct was unjustifiable and unreasonable. Yet this appeared to have been a deliberate position that the Commission took in its dealings with Mr Kelly in May 2022 in the lead up to polling day.¹⁰

(d) harm to the health of Australians;

⁹ Australian Electoral Commission v Kelly, Cth (FCA 2023).

¹⁰ Australian Electoral Commission v Kelly, Cth.

Example: Misinformation that caused people to ingest or inject bleach products to treat a viral infection.

This example is hard to argue with as this is clearly going to cause serious harm. However, it again evokes the issue of correlation and causation. A large part of the concern here is that someone would follow this advice. Maybe there are other ways of dealing with such an extreme case by making sure people are receiving the correct education regarding health. If someone's friend tells them to drink bleach, you would hope they will not follow that advice.

Another query is whether this point might be applied to mental health? If so, it is likely more information will be shut down to "keep people safe". This may be amenable to being interpreted as a means of justifying removal of posts that are merely provocative and/or offensive.

(e) harm to the Australian environment;

Example: Misinformation about water saving measures during a prolonged drought period in a major town or city.

What about a claim that wind farms do more harm to the environment than good? It is unlikely that a comment like this will cause serious harm if dismissed by the relevant people who have experience with wind farms and know better.

What about saying that wind and solar can never be the sole source of power and that coal, gas, nuclear, hydrogen, etc will always be needed?

(f) economic or financial harm to Australians, the Australian economy or a sector of the Australian economy.

Example: Disinformation by a foreign actor targeting local producers in favour of imported goods.

The Fact Sheet says that the serious harm must affect 'a significant portion' of the Australian population, economy, environment etc. We think that this is implicit in some of the listed harm factors but it is not spelt out as a requirement. For example, 'hatred against a group in Australian society'. What if the group is very small and not a significant portion of the population (a small Indigenous nation in the NT)? It is not

clear that it will be argued that the hatred, albeit directed at a specific small group, affects Australians as a whole?

- *Serious*

While 'harm' is defined, there is no definitive guidance on when it is to be regarded as 'serious'. However, clause 7 provides that in determining whether the provision of content on a digital service is reasonably likely to cause or contribute to serious harm, regard must be had to the following matters:

- (a) the circumstances in which the content is disseminated;
- (b) the subject matter of the false, misleading or deceptive information in the content;
- (c) the potential reach and speed of the dissemination;
- (d) the severity of the potential impacts of the dissemination;
- (e) the author of the information;
- (f) the purpose of the dissemination;
- (g) whether the information has been attributed to a source and, if so, the authority of the source and whether the attribution is correct;
- (h) other related false, misleading or deceptive information disseminated; and
- (i) any other relevant matter.

Some of these factors i.e. (f) and (h) also appear applicable in assessing whether the information is false, misleading, or deceptive. This raises the question of how different pieces of information will be categorised firstly to the categories of mis/disinformation and then funnelled through the conditions of causing harm or serious harm. The removal or flagging of false, misleading or deceptive content can be done via an algorithm, but the serious harm aspect would surely require human judgment. We presume this judgment will be exercised by a community standards moderator as an algorithm could not make this judgment. This is concerning because harm is in the eye of the beholder (that of the community standards moderator). How can we know whether something is really likely to cause serious harm until after the fact? The focus should be on caring for victims of serious harm, not diverting energy to trying to stop what may be out of the government's hand from occurring.

2.5 Other Issues Regarding the Bill's Application

Another cause for concern is that the Bill will cause people to seek loopholes in order to communicate mis/disinformation without being caught. Certain groups may develop new slang and code for communicating about certain topics deemed controversial. This may lead to words losing their ordinary meaning and add to the confusion mis/disinformation can cause on social media. If people communicate in code in order not to get caught, the moderation of mis/disinformation will have to become increasingly more complicated and will prove ultimately counterproductive. It may also cause different groups to become more isolated from each other and detract from the marketplace of ideas that is meant to foster healthy disagreement and discussion. If you do not understand what someone is trying to communicate, it is hard to disagree with it.

This also raises concerns about proportionality. If certain codes become widely known before private social media companies can catch on, it may mean that a view spreads quickly to some while excluding those who may not speak the language used to create the code, or have the connections with the right people to understand it.

This Bill also has the potential to misrepresent the state of the society we are in. If someone, or a group, holds a view that others find abhorrent, then they usually want to know. If social media does not truly reflect the views of the public, because the information on it is strictly filtered, then it stops us from interpreting the political or social climate clearly. Going back to the Hunter Biden example, both Democrats and Republicans would want to know the truth about what was going on in this situation to inform their vote.

The Bill also could have the opposite of the desired effect. The 'Streisand effect' is an example of how some measures taken to remove content can cause them to have increased viewership. 'In 2003, Barbra Streisand and Kenneth filed a lawsuit against amateur photographer Kenneth Adelman for violating her privacy rights. To document accelerated erosion, Adelman had uploaded aerial images of the entire California coastline to his website, with one of them showing the singer's mansion in Malibu. Beyond the fact that the courts dismissed Streisand's claim for damages and injunction, only six people had downloaded the image before the lawsuit - "including twice by her own lawyers" (Cacciottolo 2012) - whereas the website attracted over 400,000 visitors in the following month. Named after this incident, the "Streisand effect" describes the unintended consequences counter to the censor's initial

motivation for withholding information (Martin 2007).¹¹ Many accounts of misinformation may end up being seen on other platforms that are beyond the practical reach of the Bill, or through other means entirely. The misinformation could be multiplied to such a degree that a company loses the ability to effectively remove it.

Conclusion

The definitions in the bill are not easily applicable in many cases it is trying to address. The attempt to reduce harm caused by mis/disinformation is likely to cause harm itself.

Please contact the below staff if you require more information.

Dr Sharon Rodrick, Associate Director [REDACTED]

Mark Sneddon, Director [REDACTED]

Amy Bulman, Researcher

¹¹ Christian Gläsel and Katrin Paula, "Replication Data for: Sometimes Less Is More: Censorship, News Falsification, and Disapproval in 1989 East Germany" (Harvard Dataverse, 2019), doi:10.7910/DVN/AZFHYN.