Comments to the Australian Draft Social Media (Anti-Trolling) Bill

ARTICLE 19: Global Campaign for Free Expression

21 January 2022

ARTICLE 19: Global Campaign for Free Expression (ARTICLE 19) is an independent human rights organisation that works around the world to protect and promote the rights to freedom of expression and freedom of information. Over the years, ARTICLE 19 has produced a number of standard-setting documents and policy briefs based on international and comparative law and best practices on freedom of expression issues, including on the protection of reputation[1] and social media.[2] The organisation has significant experience analysing and commenting on defamation laws around the globe, from Mexico to Malaysia; and advocacy against the abuse of defamation laws to censor independent media, activists and human rights defenders.[3]

ARTICLE 19 has also been an intervener in a number of cases related to the protection of reputation and social media before the European Court of Human Rights,[4] the Inter-American Court of Human Rights,[5] the African Court on Human and Peoples' Rights - as well as to courts in national jurisdictions (e.g. most recently in Poland[6] or Indonesia[7]).

[1] ARTICLE 19, Defining Defamation: Principles on Freedom of Expression and Protection of Reputation, Updated in 2017. An earlier version of the principles were adopted by a group of highly recognised and respected experts in the area of freedom of expression and protection of reputation, and they have been endorsed by all three special international mandates dealing with freedom of expression – the UN Special Rapporteur on Freedom of Opinion and Expression, the Organisation for Security and Cooperation in Europe Representative on Freedom of the Media, and the Organisation of American States Special Rapporteur on Freedom of Expression – as well as a large number of other organisations and individuals.

[2] ARTICLE 19, Watching the watchmen Content moderation, governance, and freedom of expression, 2021, Policy Brief; ARTICLE 19, Taming Big Tech. Protecting freedom of expression through the unbundling of services, open markets, competition, and end users' empowerment, 2021, Policy Brief; ARTICLE 19, Side-stepping rights: Regulating speech by contract, 2018, Policy Brief; ARTICLE 19, Internet intermediaries: Dilemma of Liability, 2013, Policy Brief.

[3] See, among others: ARTICLE 19, Leyes del Silencio. Acoso contra la libertad de expression de expression en México y Colombia, 2021; ARTICLE 19, State of SLAPPs in Spain, November 2021, Country Report; ARTICLE 19 et al, Poland: MFRR condemns defamation lawsuit against Gazeta Wyborcza Editor-in-Chief by Polish Justice Minister, March 2021; ARTICLE 19 et al., Serbia: Lawsuits against investigative portal chills media freedom, December 2021.

[4] Among others, ARTICLE 19 intervened in the following cases: European Court of Human Rights, *Kharitonov v Russia*, no. 10795/14, judgment of 23 June 2020; European Court of Human Rights, *Delfi AS v Estonia*, no. 64669/09, judgment of 16 June 2015; European Court of Human Rights, *Mouvement Raelien Suisse v Switzerland*, no. 16354/06, judgment of 13 July, 2012; European Court of Human Rights, *Tamiz v the United Kingdom*, No. 3877/14, judgment of 19 September 2017; European Court of Human Rights, *Biancardi v Italy*, no. 77419/16, judgment of 25 November 2021.

[5] See e.g. ARTICLE 19, Amicus brief to the Inter-American Court ot Human Rights, *Tulio Alvarez v. Venezuela,* Case No. 12.663.

[6] See Public Amicus Brief by ARTICLE 19 in the District Court of Warsaw, *Telewizja Polska SA in Warsaw v Prof. Wojciech Sadurski*, Case XXIV C 276/19.

[7] See Public Amicus Brief by ARTICLE 19, Amnesty International, Cairo Institute for Human Rights, and Egyptian Initiative for Personal Rights, in the Constitutional Court of Indonesia, Judicial Review of Law No. 1/PNPS/1965 concerning the prevention of religious abuse and/or defamation; Amicus brief to the Court of Banda Aceh District,

Introduction and summary

ARTICLE 19 welcomes the opportunity to comment on the Draft Social Media (Anti-Trolling) Bill (the 'Draft Bill'). While we consider the Draft Bill represents a good faith attempt to balance the rights to freedom of expression and the right to reputation, it fails in our view to afford sufficient weight to the latter; therefore we recommend not to adopt it in its current form.

Under international human rights standards, States may, exceptionally, limit the right to freedom of expression provided that such limitations conform to the strict requirements of the three-part test[8]. This requires that limitations must be:

- **Provided for by law**. Any law or regulation must be formulated with sufficient precision to enable individuals to regulate their conduct accordingly;
- **In pursuit of a legitimate aim**, listed exhaustively as the respect of the rights or reputations of others, or the protection of national security or of public order (ordre public), or of public health or morals; and
- **Necessary and proportionate** in a democratic society, requiring that if a less intrusive measure is capable of achieving the same purpose as a more restrictive one, the less restrictive measure must be applied.[9]

Further, Article 20(2) of the ICCPR provides that any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence must be prohibited by law.

The same principles apply to electronic forms of communication or expression disseminated over the Internet.[10]

We are concerned that, in its current form, several provisions of the Draft Bill do not meet the requirements of the three-part test above. For instance, the Draft Bill establishes limits to the right to free expression regarding 'comments' posted on social media platforms, but it fails to provide a definition of what 'comment' means for the purpose of the Bill. Similarly, Section 20 imposes the obligation to establish a nominated entity in Australia for those social media companies with at least 250,000 Australian persons who hold accounts but fails to define what 'Australian' means, as well as where these persons are supposed to be based in Australia to be counted for the sake of the threshold. The lack of precise definitions on these key concepts means that the test of legality is not met, and therefore the Draft Bill is in violation of the international human rights rules and standards that protect the right to freedom of expression.

Moreover, a variety of provisions in the Draft Bill do not appear to respect the principle of proportionality. For instance, to attribute to social media platforms the publishers' liability with

---

Attorney General's Office of the Republic of Indonesia vs. Dr. Saiful Mahdi M S.Si, M.Sc Bin (Alm) Abdullah, Case No 432/PID.SUS/2019/PN BNA.

[8] See: UN General Assembly, International Covenant on Civil and Political Rights, 16 December 1966, UN Treaty Series, vol. 999, p. 171, Article 19(3).

[9] HR Committee, *Belichkin v. Belarus*, Communication No. 1022/2001, UN Doc. CCPR/C/85/D/1022/2001 (2005).

[10] General Comment No. 34 12 September 2011, CCPR/C/GC/34, para 43. The General Comment states that "any restrictions on the operation of websites, blogs or any other internet-based, electronic or other such information dissemination system, including systems to support such communication, such as internet service providers or search engines, are only permissible to the extent that they are compatible with paragraph 3. Permissible restrictions generally should be content-specific; generic bans on the operation of certain sites and systems are not compatible with paragraph 3. It is also inconsistent with paragraph 3 to prohibit a site or an information dissemination system from publishing material solely on the basis that it may be critical of the government or the political social system espoused by the government.

regards to comments posted by end-users is a disproportionate provision, which fails to consider the numerous differences between the role and activity of publishers and those of social media companies. Similarly, the imposition, on social media platforms, of the duty to collect personal data (at least the name, email address and phone number) of each user is a disproportionate interference with the user's rights: a less intrusive solution should have been chosen to guarantee the right of the reputation of individuals.

In the following paragraphs, we provide a more detailed analysis of the Draft Bill's provisions which we consider to be in violation of international human rights law and standards on freedom of expression.

1. Limiting immunity from liability will have a chilling effect on freedom of expression

ARTICLE 19 is concerned that if adopted in its current form, the Draft Bill will have the effect to strongly limit the immunity from liability of social media platforms for the content posted by users, in violation of international standards on freedom of expression.

Indeed, Part 2 of the proposed Bill attributes social media companies the qualification of publishers with regards to that content. In turn, this implies that social media companies can be held liable for defamatory comments posted by users and they can only defend themselves from this liability by setting the compliance system described in Section 16. With regards to this system of liability, we note the following.

We have repeatedly claimed that unduly limiting immunity from liability for content posted by users produces a number of dangerous trade-offs for free expression. It gives social media companies an incentive to either filter and remove as much of users' content as possible or to be entirely neutral and not remove any content at all. In other words, it would either lead to increased censorship or remove any incentives for companies to engage in content moderation. In our view, both these outcomes would be highly undesirable for the protection of freedom of expression online. For this reason, we believe that social media companies must continue to benefit from broad, or at least conditional, immunity from liability, and we express concerns that the Bill will create incentives for social media platforms to monitor and over censor users' content.

Our call is consistent with the provisions in the Joint Declaration on Freedom of Expression and the Internet, which requires that '*No one who simply provides technical Internet services such as providing access, or searching for, or transmission or caching of information, should be liable for content generated by others, which is disseminated using those services, as long as they do not specifically intervene in that content or refuse to obey a court order to remove that content, where they have the capacity to do so ('mere conduit principle')'.*[11]

Moreover, successive UN Special Rapporteurs on the promotion and protection of the right to freedom of opinion and expression (Special Rapporteurs on FoE) have stated several times that censorship should never be delegated to a private entity and that States should not use or force intermediaries to undertake censorship on its behalf.[12] Further, former

[11] See: Joint Declaration on Freedom of Expression and the Internet adopted on 1 June 2011 by The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, para. 2(a).
[12] The Report of the Special Rapporteur on FoE, 16 May 2011, A/HRC/17/27, para 43.

Special Rapporteur David Kaye posited that "*smart regulation, not heavy-handed viewpoint-based regulation, should be the norm, focused on ensuring company transparency and remediation*."[13] Although he did not rule out the possibility of regulation under certain conditions, he reiterated that States should only seek to restrict content pursuant to an order by an independent and impartial judicial authority, and in accordance with due process and standards of legality, necessity and legitimacy.[14] The Special Rapporteur also noted that States should refrain from imposing disproportionate sanctions, whether heavy fines or imprisonment, on Internet intermediaries, given their significant chilling effect on freedom of expression.[15] His successor and the current Special Rapporteur on FoE, Irene Khan, has come to a similar conclusion.[16]

Other international human rights standards, such as the Manila Principles on intermediary liability, also recognise the need to shield intermediaries from liability for third-party content.[17]

Besides, the likely chilling effect of liability on freedom of expression has been recognised by the same Australian Government, which in its Explanatory Paper accompanying the Bill declares:

'Liability concerns may have a chilling effect on free speech, as people who administer social media accounts may censor comments or disable functionalities due to a fear of being held liable in defamation for content that they did not post.

In some cases, the Voller decision may have contributed to decisions to limit the ability for the general public to interact with social media posts about news and current events.'

Indeed, the same concerns that the Australian Government expresses with regards to people who administer social media accounts can be expanded to social media platforms. Imposing liability on social media platforms would indeed incentivise over-removal of content and other measures that limit users' freedom of expression. What is more, in the case of social media platforms, this problem will present at a substantially bigger scale, and thus the negative impact on free speech will be substantially higher.

2. The defence provided by Sections 15 and 16 does not adequately safeguard freedom of expression

i. Social media companies are not publishers *tout court*

Section 14 of the Bill establishes that social media platforms are to be considered publishers with regards to the comments posted by users. It also clarifies that, with regards to defamation claims, social media platforms cannot use the defence of innocent dissemination but can only defend themselves if they establish the system provided for in Sections 15 and 16.

First, ARTICLE 19 has long defended that social media platforms cannot *tout court* be compared to publishers, like newspapers or media companies. Although it might be said that some social media platforms are increasingly behaving like traditional incumbent media

[13] Special Rapporteur on FoE, Report on content regulation, 06 April 2018, A/HRC/38/35, para 66.
[14] *Ibid.*
[15] *Ibid.*
[16] Special Rapporteur on FOE, Report on disinformation, 13 April 2021, A/HRC/ 47/25, para 91.
[17] See Manila Principles on Intermediary Liability, March 2015. The Principles have been endorsed by over 50 organisations and over 100 individuals.

companies, some key differences remain. In particular, they have millions of users posting content on their platform. Users are the primary publishers of this content. Social media platforms select and organise that content. As such, they exercise a form of editorial responsibility – a process sometimes referred to as 'content curation'. However, they are not responsible for the content itself unless they modify it sufficiently that it might be said to be their own[18].

Based on the above, we consider this *fictio iuris* an overstep, and we warn against the impact it will produce on freedom of expression. Furthermore, we believe that this provision does not respect the test of proportionality that is required under international freedom of expression standards.

Finally, we note that the Draft Bill fails to define what should fall within the category of 'comments', which will trigger the liability. The term 'comment' has a specific meaning in the international standards with regards to defamation, which distinguish between comments on the one hand, and statements of fact on the other, including for the common law defence of fair comment. Therefore, this lack of definition creates legal uncertainty to the detriment of all actors involved.

ii. The collection and communication of personal data does not comply with privacy and freedom of expression standards

The Complaints scheme described in Section 16 raises additional challenges with regards to the right to privacy in online communications and the right to anonymity. The scheme obliges social media platforms to collect and retain the name, phone number, and email address of every Australia-based 'commenter'. If a complainant requests the social media platform to disclose the relevant content details of the commenter, the social media platform shall: (i) disclose the country location data of the commenter and (ii) if the complainant is dissatisfied with the way the complaint has been handled by the social media platform, it must disclose the relevant contact details of the commenter to the complainant in order to assist the complainant with the potential institution by the complainant of a defamation proceeding against the commenter in relation to the comment. Before doing so, the social media platform needs to seek the consent of the commenter. If this is denied, the claimant can seek an end-user disclosure order from the court, following which the social media platform will be obliged to disclose the commenter's personal details.

First, it is noted that Section 16(1) provides that in order to make a complaint, and thus to require the commenter's details, the applicant shall have a reason to believe that they have the right to obtain relief against the commenter in a defamation proceeding that relates to a commenter's comment posted on a page. From the wording of the Bill, the admissibility is based on the vague 'reason to believe' test on the part of the complainant. As known, defamation cases imply a fine balancing of the conflicting rights involved: the right to reputation of the complainant and the right to freedom of expression of the defender. Such a balancing exercise should be carefully conducted according to relevant international standards, and it cannot be left in the hands of the claimant. Indeed, defamation claims are widely used to suppress expression and to censor voices and opposition. To expose commenters to the possibility of their personal data being shared with complainants based on a simple 'reason to believe' test is a disproportionate exposure, which will certainly have

[18] See Watching the watchmen: Content moderation, governance, and freedom of expression policy *op. cit*.

a chilling effect on freedom of expression and does not respect the proportionality test provided under international human rights law.

What is more, as reminded by the UN Special Rapporteur on FoE, to guarantee the right to privacy in online communications is essential to ensure that individuals have the confidence to freely exercise their right to freedom of expression. The inability to communicate privately substantially affects individuals' freedom of expression rights. As such, the Special Rapporteur has expressed concerns over States and private actors monitoring and collecting information about individuals' communications and activities on the Internet. These practices can constitute a violation of Internet users' right to privacy, and ultimately impede the free flow of information and ideas online[19]. The Special Rapporteur also recommended that States should ensure individuals can express themselves anonymously online and refrain from adopting real-name registration systems[20]. Further, he recommended that States refrain from making the identification of users a pre-condition for access to digital communications and online services.[21]

The Draft Bill completely ignores these international standards. Indeed, in order to comply with Section 16 of the Bill, social media companies will be forced to ask end-users the information they are supposed to share in case of complaints, and in particular the name, email address, phone number and other details as specified in the legislative rules (Section 6). It is evident that this provision is incompatible with the right of users to express themselves anonymously online. It is also incompatible with the Australian Privacy Principle 2, which provides that individuals must have the option of not identifying themselves, or of using a pseudonym, when dealing with a regulated entity.

Moreover, the Draft Bill requires social media platforms to geolocalise end-users and to disclose their locations to the complainants when a defamation complaint is filed. In order to comply with this requirement, social media platforms will need to use by default geolocation technologies with regards to end-users. This raises major privacy challenges and threatens the use of privacy-protecting tools designed to anonymise user information or IP addresses. As recalled by the UN Special Rapporteur, though, the right to anonymity covers not only the content of communications but also identifying factors (the metadata) like the IP address etc. Users seeking to ensure full anonymity may use tools such as virtual private networks (VPNs), proxy services, anonymising networks and software, and peer-to-peer networks[22]. The location disclosure requirement in the Draft Bill will evidently put this at risk.

To be able to comply with the Draft Bill's requirement, social media will need not only to collect the data, but also to store them, and to provide access to their nominated entity in the country (for a further discussion of nominated entities, see below). The storage of personal data raises a variety of challenges and exposes individuals to a variety of risks. Indeed, the Draft Bill says nothing about where the data have to be stored, but only provides that the nominated entity shall have access. This implies that end users are left with no protection against security breaches, or against misuses or abuse of their personal data. Moreover, they have no right to remedy for breaches, nor right to transparency about access' requests and the rights to rectification and to be forgotten. In conclusion, the Draft Bill imposes a vast collection of personal data from individuals, but provides no guarantee that the data will be gathered, stored and used according to international privacy and data protection standards.

---

[19], Report of the Special Rapporteur May 2011, *op. cit*., para 53

[20] *Ibid.* para 84

[21] See also the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Encryption and anonymity follow-up report, Research Paper 1/2018, June 2018.

[22] See: UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Report on Encryption, anonymity, and the human rights framework, 22 May 2015, A/HRC/29/32 para. 9.

As such, the Draft Bill incentivises social media platforms to perform high risk activities in a regulatory vacuum, while it should at minimum set a system of adequate guarantees for the protection of individuals' personal data. We finally note that it is of utmost importance that the discussions around the Draft Bill adopt a holistic approach, which includes also the Australian Government's plans to review the Privacy Act and to issue a Privacy Enhancing Legislation that focuses on social media and other online platforms.

3. The requirement to have a nominated entity in Australia raises human rights and competition concerns

Section 20 of the Draft Bill establishes that if the provider of a social media service is a foreign body corporate and the service has at least 250,000 Australian account-holders (or the service is specified in the legislative rules), the provider of the social media service must have a nominated entity in Australia, which shall have access to the end-users' data required by the Draft Bill and be able to manage the complaints system.

First, ARTICLE 19 notes that we have been criticising similar obligations of having a nominated entity in the country elsewhere. Such obligations have been primarily introduced in countries with the absence of due process and an independent judiciary, such as Turkey, Russia or Nigeria most recently. In our experience, these requirements risk becoming the long arm of the law enforcement regimes and eventually become complicit in human rights abuses in the country as companies will not be able to resist the demands of law enforcement authorities for long. We are concerned that the worst offenders in terms of Internet censorship in the region and beyond will refer to the Australian legislation in order to justify restrictions on freedom of expression and expansion of their efforts to stamp out dissent. Such requirements can backfire.

Second, the threshold of 250,000 Australian account-holders is problematic. First, the Draft Bill refers the number to 'Australian persons', without specifying: if this includes citizens only, or also other people; if the 'Australian persons' have to be in Australia to be included in the threshold.

Moreover, the threshold of 250,000 end-users is relatively low and might include small social media platforms. In this scenario, the small players will be strongly disadvantaged and suffer the regulatory burden imposed by the Draft Bill way more than the bigger platforms. As such, the Draft Bill will have as a trade-off to entrench the dominant position on the market of the bigger platforms in an already extremely concentrated market. As we have highlighted on various occasions, this high concentration on the social media market is detrimental to users' right to freedom of expression, and States should introduce measures that dilute this concentration, not entrench it.[23]

[23] See: ARTICLE 19, Taming Big Tech: Protecting freedom of expression through the unbundling of services, *op. cit.*