



Who cares about free speech?

Attitudes towards freedom of expression and their implications for content moderation on social media

Findings from a global survey



Introduction

Across the globe, governments are scrambling to regulate content on social media to limit various (contentious) categories such as disinformation and hate speech. In many cases, measures include so-called intermediary liability laws that oblige platforms to remove illegal or even “harmful” content once they are notified. In Europe, this trend was kick-started by Germany’s Network Enforcement Act (NetzDG), which stipulates that social media companies with over 2 million users must remove manifestly illegal content within 24 hours after notice or risk hefty fines. Justitia documented how the NetzDG became the blueprint for more than 20 such restrictive regulatory proposals or laws worldwide – in both authoritarian states and liberal democracies.¹ Regulating online content is also a hot-button issue at the EU level. In December 2020, the European Commission presented its proposal for the Digital Services Act (DSA)², which introduced new obligations for intermediary service providers, including special obligations for “very large online platforms”.³ These include submitting an annual assessment of significant risks stemming from their functions (Article 26), implementing mitigation measures for such risks (Article 27), submitting independent audits to assess compliance (at their own expense) (Article 28) and, ‘upon a reasoned request’, providing the EU with relevant data necessary to assess compliance (Article 31). The Commission may impose fines in the event of non-compliance (Article 59).

Even if platforms were to comply with all requirements set out by the DSA, content moderation practices that reflect national or regional norms are bound to conflict with the global nature of social media platforms. Platforms cater to billions of users across nations and cultures with varying definitions of and tolerance for offensive content.

To understand and conceptualize the varying standards for free speech among different governments and peoples, Justitia conducted a survey in 2021 to study the support for free speech around the world. The survey included approximately 50,000 respondents in 33 countries, including 3 Northern European countries: Sweden, Denmark, and Norway. This short paper will provide an insight into two aspects of the extensive survey – namely, attitudes toward free speech when it comes

¹ Jacob Mchangama & Joelle Fiss ‘The Digital Berlin Wall: How Germany (Accidentally Created a Prototype for Global Online Censorship’ (2019) *Justitia* http://justitia-int.org/wp-content/uploads/2019/11/Analyse_The-Digital-Berlin-Wall-How-Germany-Accidentally-Created-a-Prototype-for-Global-Online-Censorship.pdf>; Jacob Mchangama & Natalie Alkiviadou ‘The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship – Act Two’ (2020) *Justitia* <https://justitia-int.org/wp-content/uploads/2020/09/Analyse_Cross-fertilizing-Online-Censorship-The-Global-Impact-of-Germanys-Network-Enforcement-Act-Part-two_Final-1.pdf>

² <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020PC0825&from=en>

³ Those with over 45 million users in the EU.



to offensive speech and disinformation and respondents’ preferences with respect to social media regulation. The paper will then use these findings to briefly sketch a number of implications for content moderation at scale.

Attitudes toward Free Speech

To assess the **actual** support for free speech in a country, the survey includes a composite measure, the **Justitia Free Speech Index**, based on answers to eight “tough” questions. When we match the Justitia Free Speech Index scores with country scores from V-Dem’s Freedom of Expression Index, there is a clear, positive association. This means that public opinion about free speech tends to go hand-in-hand with the actual enjoyment of this right.

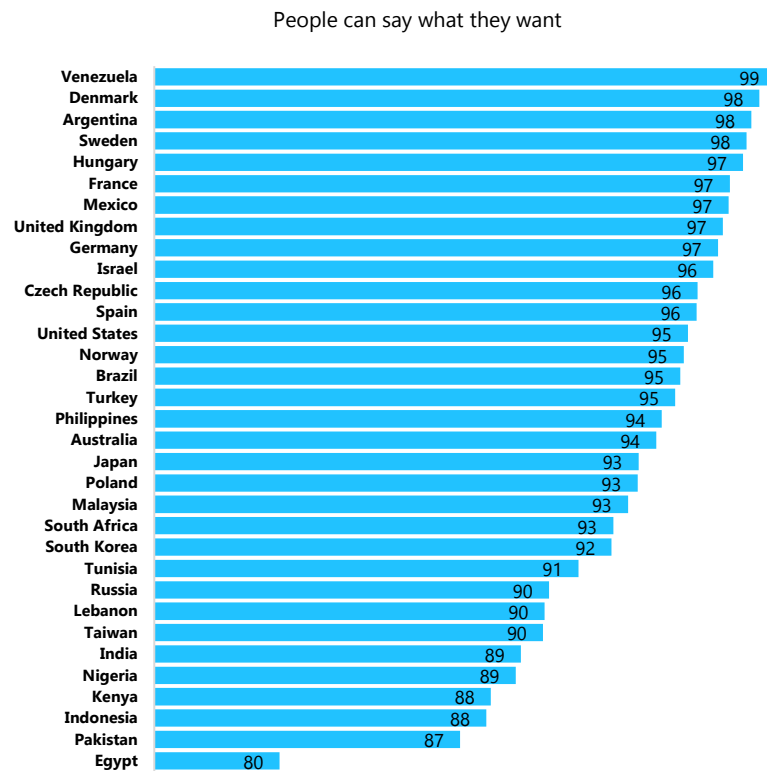
Figure 1: Global variation in the Justitia Free Speech Index





There is strong universal support for the right to free speech in the abstract. When respondents were asked about the importance of being able to exercise this right, 94% of people across the spectrum of all 33 countries agreed that it is important or very important that “people can say what they want”.

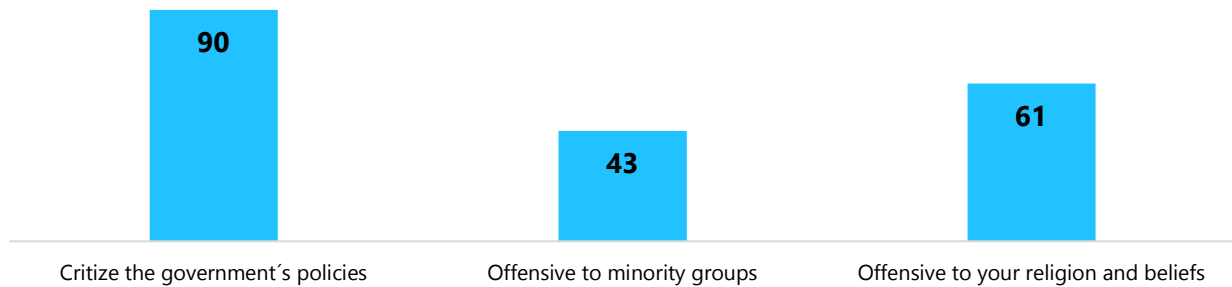
Figure 2: Support for freedom of expression without censorship in individual countries



There is also strong support for the right to criticize the government. The median support across countries is 90% (in comparison to 43% and 61% for offending minority groups and offending religion/belief, respectively). There is, thus, greater support for the right to criticize government in comparison to support for offensive statements. Only in 6 countries is the variation in relation to the acceptability of speech across the spectrum (to criticize the government and be offensive) relatively small. As noted in the survey’s report, this shows that only a handful of countries reflect the belief that free speech should apply broadly, no matter the “target”. As a result, free speech is endorsed in the abstract (as reflected above). However, free speech extends to ideas that we may not like, and this aspect received a lower level of acceptance in most countries.



Figure 3: Median support for particular types of free speech across all countries

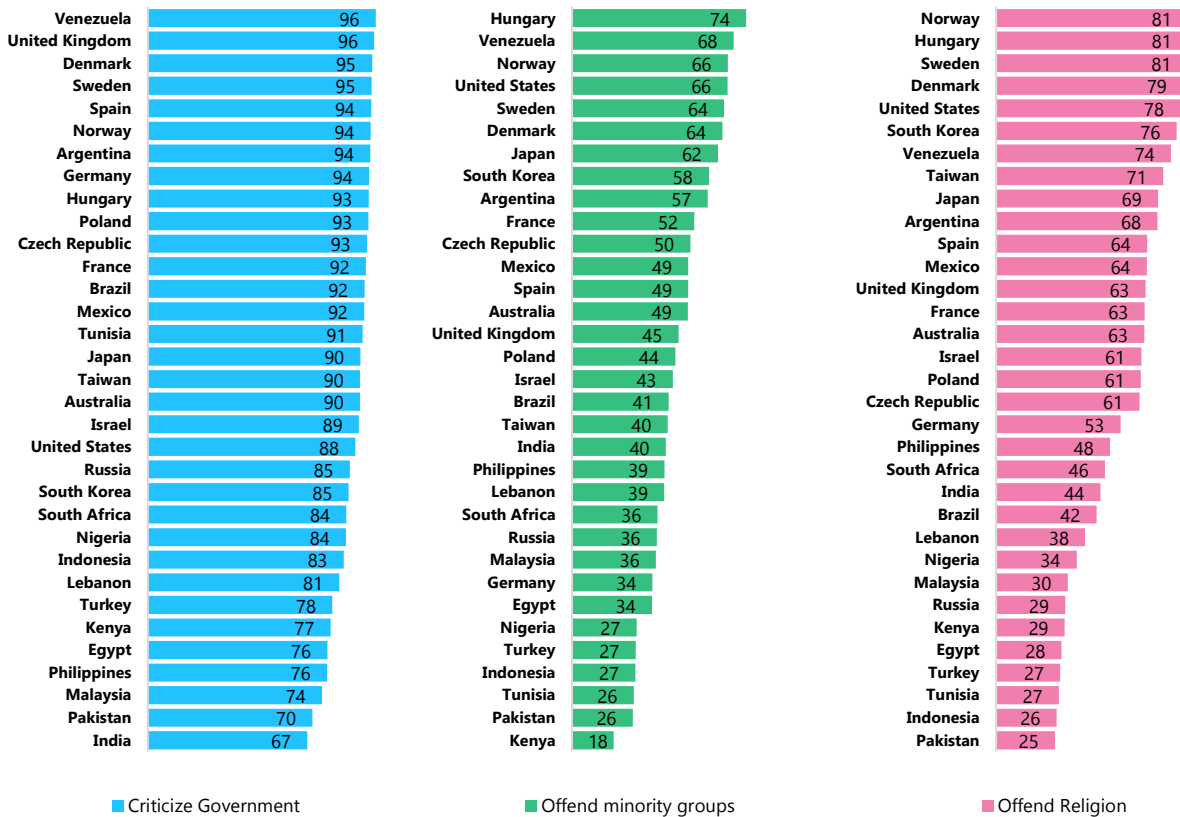


However, when confronted with controversial statements – e.g., statements offensive to minorities or religions, statements supportive of homosexual relationships or insults to the national flag, support is generally lower and varies much more between countries and across issues and individuals. Likewise, variations between countries increase and the rankings differ when people are confronted with potential trade-offs regarding information that might be sensitive to national security, harm economic stability or undermine the handling of epidemics. Respondents were asked whether speech offensive to minority groups and statements offensive to their own religion or beliefs should be permissible. Across countries, there is less than 50% support for free speech with respect to statements offensive to minority groups in more than half (22/33) countries. In almost half the countries (15/33), there is less than 60% support for free speech with respect to offence to religion or belief. As such, there is a demand in several countries to restrict speech that may be deemed, for example, offensive or blasphemous.

However, there is no universal agreement on whether statements offensive to minorities or religions should be tolerated. National figures reveal this. In the Northern European countries included in this survey, 65% of the population believes that free speech should extend to statements offensive to minority groups while around 80% thinks that statements offensive to religion should be allowed. In Kenya, Indonesia, Turkey and Tunisia, only between 18% and 27% of the population favors tolerating statements offensive to minorities, while 26%-29% of the population in those countries favors tolerating expressions offensive to religion. However, even within European democracies, attitudes towards freedom of expression and its limits vary. Northern European states are generally more supportive of robust protections for freedom of expression than others – including, most significantly, large, influential countries like France and Germany. Germany, in particular, is an outlier for Europe since its support for speech that is offensive to minority groups falls to 34% (just under Malaysia and on the same metric as Egypt) and 53% for religious offence (“revilement of religious faiths” is punishable under German law), which is just above the Philippines. Germany’s history of minority persecution under Nazi rule may have contributed to the formation of public opinion on such matters.



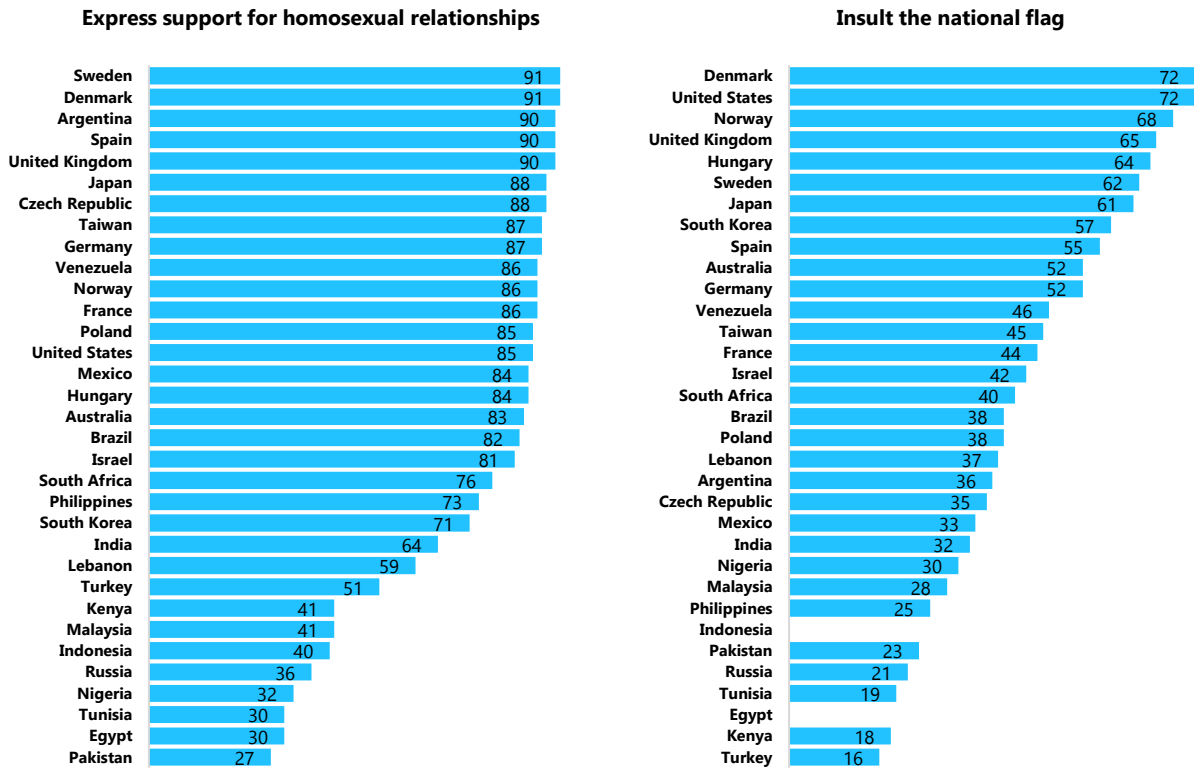
Figure 4: Support for ability to criticize government, offend minority groups, or offend religion



In two additional questions about controversial issues, we have asked whether people should be allowed to express support for homosexual relationships or insult the national flag. The responses to these questions also vary significantly between nations. In Denmark, Sweden, Spain, the UK, and Argentina, there is almost uniform support for allowing people to express support for homosexual relationships. By contrast, there is widespread opposition to allowing such statements in Kenya, Russia, and all Muslim-majority countries in our sample.



Figure 5: Allowing statements that support homosexual relationships or insult the national flag

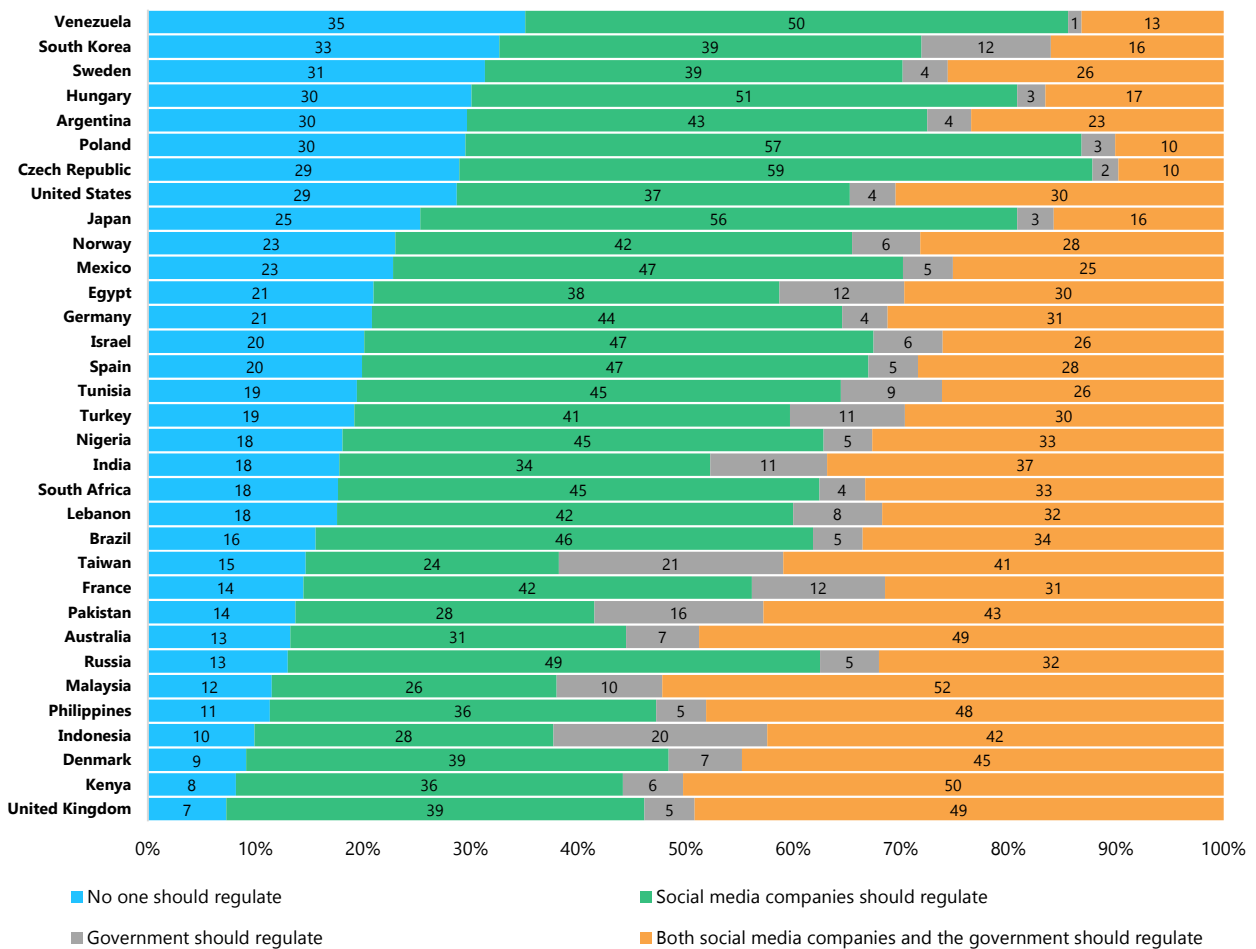


Attitudes toward the Regulation of Social Media

Given the conflicting attitudes about *where* the line should be drawn with respect to social media, the question of *who should* draw the line is crucial. In all the countries surveyed by Justitia, a majority prefers some kind of regulation of social media content. However, only a few want the government to take sole responsibility for this. People in two-thirds of the countries surveyed prefer such regulation to be carried out by social media companies themselves while a plurality of the remaining countries prefers the responsibility for regulating content to rest with social media companies alongside national governments. Within the Northern European countries studied in the report, there appeared to be broad support for the proposition that social media companies should be responsible for regulating content: Norway (42%), Sweden (39%) and Denmark (39%). That said, some respondents also accepted joint regulation by companies and governments together – Norway (28%), Sweden (26%) – while Danish results show a preference for this type of regulation (45%).



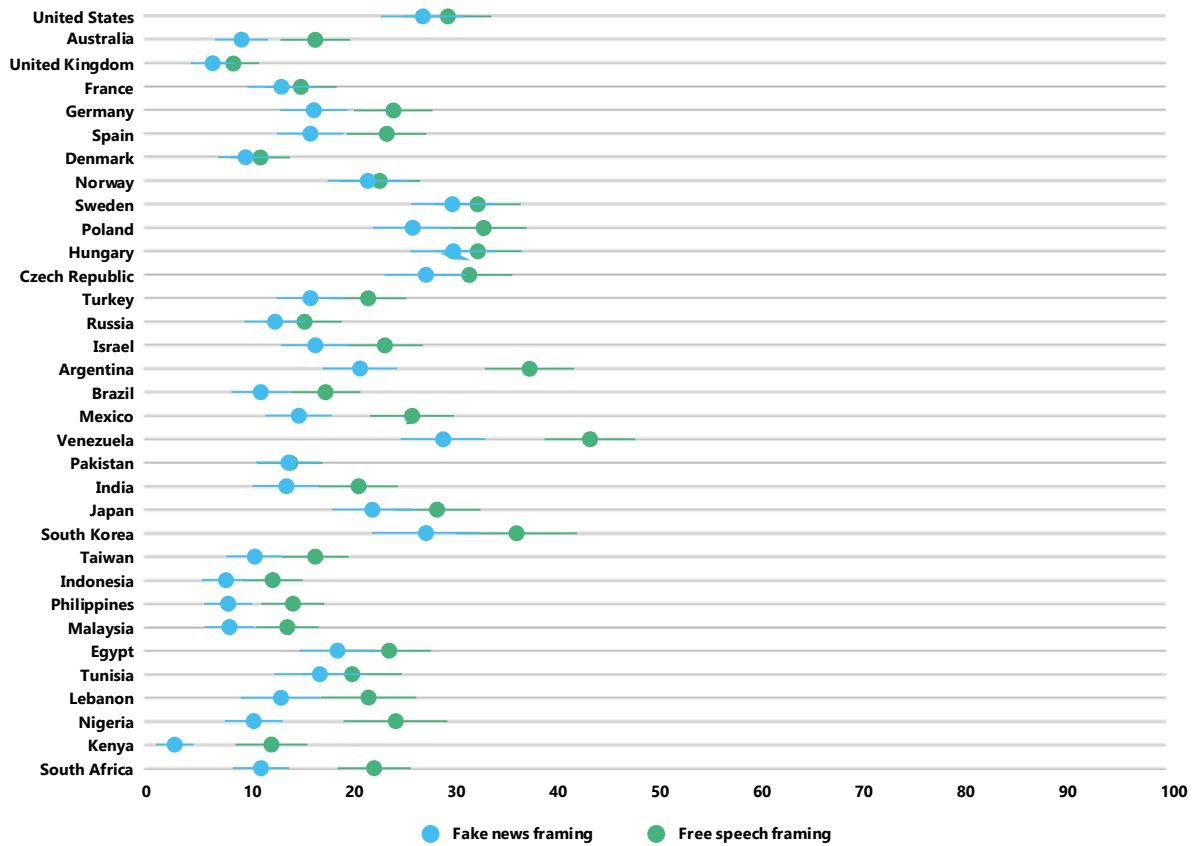
Figure 6: Support of regulation of social media content



Interestingly, the way the issues are framed may change people’s perceptions of regulation. All nations showed higher levels of support for no regulation when the issue was presented as a “repression of free speech”. When framed as “regulating disinformation”, support for no regulation dwindled. This points to a need for a balanced and fair discussion of regulation that takes into consideration the issue of disinformation (and other kinds of harmful speech) as well as the impact of regulation on the practice of free speech.



Figure 7: Sensitivity of support for no regulation of social media content (by framing)





Concluding Comments

While support for the ideal of free speech in the abstract is evident, it is clear that there is no universal agreement on how to define freedom of expression or its scope. Online content moderation of contentious areas of speech is complicated by (i) the sheer number of social media users and their different backgrounds and belief systems, which lead to (ii) a lack of global agreement as to what should and should not be allowed and (iii) the creation of a situation in which content moderation standards and practices cannot possibly satisfy all users and governments. As a result, meaningful content moderation at the scale of global platforms aiming to shield users from offense and (perceived) harm while simultaneously preserving the fundamental right to freedom of expression does not appear to be feasible unless one of these principles is subjugated to the other, which – in either case – is likely to alienate significant numbers of users in various parts of the world.

That said, the survey revealed that most users feel more comfortable if the companies themselves deal with speech regulation, while there is strong scepticism of governments regulating social media content alone. While addressed to and solely legally binding on states, International human rights law (IHRL) includes a global set of norms which could provide platforms a set of generally agreed upon principles to guide their standards and practices of content moderation, or what *Justitia* has termed a “Framework of First Reference”.⁴ The appropriateness of IHRL for this purpose is also the position endorsed by the former and present Special Rapporteurs on the Freedom of Opinion and Expression, David Kaye and Irene Khan. Kaye noted that social media companies have a “geographically and culturally diverse user base” and proposed IHRL as a framework for “holding both States and companies accountable to users across national borders”.⁵

⁴ “A Framework of First Reference: Decoding a Human Rights Based Approach to Content Moderation in the Era of Platformization” *Justitia* (Forthcoming 2021)

⁵ Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018) A/HRC/38/35, pg.14