CELE
Centro de Estudios en Libertad de Expresión

Documento de posición

32

CELE submission to the Committee of Ministers' public consultation on the draft recommendation on online safety and empowerment of content creators and users

Nicolás Zara

7 de agosto de 2025

Nicolás Zara, "CELE submission to the Committee of Ministers' public consultation on the draft recommendation on online safety and empowerment of content creators and users", Documento de posición No. 32 (ENG), Centro de Estudios en Libertad de Expresión (CELE), Buenos Aires (2025)

UP Universidad de Palermo

Facultad de Derecho

Centro de Estudios en Libertad de Expresión

CELE submission to the Committee of Ministers' public consultation on the draft recommendation on online safety and empowerment of content creators and users

Nicolás Zara - Researcher

cele@palermo.edu

Buenos Aires, Argentina

August 7, 2025

Introduction

The Center for Studies on Freedom of Expression and Access to Information (CELE) is an academic research center affiliated with Universidad de Palermo in Argentina. The Center provides technical and legal analysis on issues affecting this fundamental right, and since 2012 has been studying freedom of expression on the Internet as a specific research area. The Center is a leading voice on the promotion and protection of freedom of expression nationally, regionally, and internationally.

This document was prepared in response to the Public consultation on the draft Committee of Ministers Recommendation on Online Safety and Empowerment of Content Creators and Users[1] (hereinafter, the Recommendation). It aims to contribute to the Council's efforts to develop a framework that ensures a free, open, and accessible internet for all, while also creating an enabling environment for the online exercise of the right to freedom of expression and other rights.

[1] Council of Europe, Public consultation on the draft Committee of Ministers recommendation on online safety and empowerment of content creators and users. Available at: https://www.coe.int/en/web/freedom-expression/-/public-consultation-on-the-draft-committee-of-ministers-recommendation-on-online-safety-and-empowerment-of-content-creators-and-users

On the utilization of online safety as a legal category

1. The Recommendation highlights the growing concern over "risks to the online safety of users and content creators"[2] that come with the expansion of the online environment and which "can result in harm to individuals, groups, society, or societal interests such as democracy, the rule of law, and the free flow of information"[3]. It stresses the need to protect the Freedom of Expression of users that share content as well as "the safety of users at risk of being silenced by content carrying the risk of harm"[4].

2. Online Safety is now a commonplace phrase in internet-related discussions. Many states, such as the United Kingdom[5], Sri Lanka[6], Australia[7], and Ireland[8], have adopted "Online Safety" acts. However, rarely do these laws define what online safety is[9] and the content of these legislations varies. Moreover, while some of these establish co-regulatory schemes, others enable local authorities to issue codes that flesh out the duties set out by the law and guide the internet platform's practices.
3. As the Recommendation acknowledges, "the online environment has become one of the principal means for the exercise of the right to freedom of expression and information"[10]. Any regulation by the state to exert control over internet platforms can have a profound impact on the freedom of the people to receive and impart information. It is therefore relevant to remember that online safety is not among those interests considered legitimate for restricting this right under the European Convention on Human Rights[11]. Under the case law of the European Court of Human Rights, the concept of "public safety" in the context of Article 10.2 must be interpreted narrowly[12]. We suggest that references to "online safety" be removed from the Recommendation, since it is not an adequate legal basis for action. Instead, we propose that they are reframed in the language of rights.

---

[2] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par.1.

[3] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par.1.

[4] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. m.

[5] Online Safety Act 2023 (c. 50)

[6] Online Safety Act, No. 9 of 2024

[7] Online Safety Act, No. 76 of 2021

[8] Online Safety and Media Regulation Act 2022, No. 41 of 2022

[9] With the exception of the Australian law, which defines "online safety for Australians" as "the capacity of Australians to use social media services and electronic services in a safe manner"

[10] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. g.

[11] European Convention on Human Rights, Article 10.2

[12] European Court of Human Rights, Stoll v. Switzerland [GC], par. 54; Gormus and Others v. Turkey, par. 37.

4. According to the definition provided by the Recommendation, "online safety implies an enabling online environment where users can communicate, access information, and participate in online activities without being unduly exposed to risks of violence, exploitation, discrimination, and other unlawful interferences with their human rights"[13]. It further explains that "enhancing online safety contributes to creating an enabling online environment for the enjoyment of human rights and specifically freedom of expression, in which users and content creators can freely participate and engage in public debates and have access to trustworthy sources of information and a variety of points of views, while retaining control over their online experience in accordance with their preferences".[14]
5. This definition of "online safety" can be easily reconstructed in the language of rights: the interest the Recommendation pursues is the protection of the rights of freedom of expression and privacy of users. This should be made clear in its text. Here's why:
6. First, "online safety" is a vague and usually poorly defined [15] –if defined at all– concept that lends itself to confusion, particularly because it has been adopted with different scopes across the various pieces of legislation that have chosen it as a framework. In contrast, Human Rights Law offers a shared language and a broadly accepted set of principles across nations. Basing the Recommendation on this foundation provides comparative advantages in terms of foreseeability and fairness.[16]
7. When the law incorporates terms like "online safety" and "user safety", it is adopting the language used by internet service providers in their operations, within their "trust and safety" function, that is, the *"practices employed by digital services to manage content and conduct related risks to users and others, mitigate online or other forms of technology-facilitated abuse, advocate for user rights, and protect brand safety"*.[17]
8. Naturally, each platform has a distinctive conception of what safety means, and it is necessarily tailored to their business needs. By adopting the terminology of the private sector, legislation accepts playing by their rules and consolidating the *status quo*. Instead of having the difficult, much-needed conversations on freedom of expression, its paradoxes, and its acceptable limits in a democratic society, the legal adoption of

[13] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 2.

[14] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 2.

[15] See, for instance, HM Government, Internet Safety Strategy – Green paper, October 2017, p. 7: "[Online safety] covers a diverse range of issues from online trolling and hate speech to location-sharing within social media platforms"

[16] Agustina Del Campo, Nicolás Zara and Ramiro Álvarez Ugarte, Reclaiming Human Rights for Platform Governance: Proposals for Restoring Their Centrality in the Era of Risks, Proceedings of Fourth European Workshop on Algorithmic Fairness, PMLR 294:265-280, 2025, p. 3.

[17] Digital Trust and Safety Partnership, Trust & Safety Glossary of Terms, retrieved from https://dtspartnership.org/glossary/, p. 11.

terms like "online safety" outsources these hard questions to the same corporations they deem to regulate.

9. As a result, framing the problem in terms of "online safety" narrows the focus to digital aspects, even when many of the underlying issues are broader in nature. Furthermore, as a term rooted in corporate discourse, it tends to highlight corporate concerns and thus invites corporate-driven solutions. Legislation, however, must adopt a wider perspective—one that goes beyond the interests or responsibilities of a single actor within the ecosystem.
10. Most importantly, the respect for the right of others constitutes an acceptable basis for restrictions under Article 10 of the Convention.
11. The "Principles for Platforms" section of the Recommendation has "online safety" as one of its central concerns[18]. It must not be overlooked that most digital intermediaries' conception of "safety" "far exceeds the legitimate objectives contemplated to restrict freedom of expression, association and assembly than Human Rights treaties do"[19], as they wish to keep as many users as possible within their platforms, which is easier without some of the most shocking or disturbing expressions protected under freedom of expression.

Risks and rights:

12. The Recommendation aims to guide states on the adoption of frameworks "that address risks to online safety, mitigate risks of harm and foster an enabling online environment that promotes the enjoyment of human rights online"[20]. It focuses on "how to address online risks that result from or impact on the exercise of the right to freedom of expression"[21]. Such risks include:
    a. *"Risks to personal and community safety and well-being associated with content and communication that users may be exposed to, interact with, or be targeted by;*
    b. *Risks to the democratic process, information integrity, and informed public discourse;*

[18] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 23-35.

[19] Agustina Del Campo and Ramiro Alvarez-Ugarte. (2025). Trust and Safety's Blindspots: A Latin American Perspective, p. 10. In Maia Levy Daniel, Amanda Menking, Marlyn Thomas Savio, & Jean Claffey, Trust, Safety, and the Internet We Share: Multistakeholder Insights, Abingdon, United Kingdom: Taylor and Francis (forthcoming 2025). Preprint retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5355934.

[20] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 6.

[21] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 6.

c. *Risks associated with the systems deployed by providers which may interfere with the rights to freedom of expression, privacy and personal data protection, and other rights."*

13. The Recommendation is clear elsewhere[22] in stating that any mitigation measure must be not only necessary and proportionate but also comply with the principle of legality: it must be precise in its wording and foreseeable in its effects[23]. However, the list of possible risks is not exhaustive, and most of them are not clearly stated. Therefore, it is of little use for states to use it in their internal lawmaking processes. A broad formulation of potential undesirable situations that users can encounter is not enough to justify state interventions. If it were to be replicated by internal legislation, it would most likely not comply with the legality principle. As a result, we suggest that it be removed.
14. The inclusion of "risks to information integrity" strikes us as particularly worrisome. While information integrity "is not a fixed analytical framework, and still lacks a clear definition"[24], it has been adopted in recent years by the United Nations and its affiliated institutions to address the impacts of digital technologies on the communication ecosystem. While recognizing that the mass adoption of digital technologies has created opportunities for "connecting individuals and communities on a previously unthinkable scale and presenting unparalleled opportunities for the diffusion of knowledge, cultural enrichment and sustainable development",[25] the Information Integrity agenda is concerned with how that same technology has given an outlet for different forms of speech that it considers disquieting although not necessarily illegal.
15. In the context of the institutions of the European Union, the Second report on FIMI[26] by the European External Action Services (SG.STRAT) follows the definition of information integrity given by the Global Declaration on Information Integrity Online, which conceptualizes it as "an information ecosystem that produces accurate,

---

[22] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 38.

[23] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. s.

[24] EPRS | European Parliamentary Research Service, Information integrity online and the European democracy shield, December 2024, p. 2., available at https://www.europarl.europa.eu/RegData/etudes/BRIE/2024/767153/EPRS_BRI(2024)767153_EN.pdf

[25] United Nations. *Global Principles for Information Integrity.* New York: United Nations, 2024. https://www.un.org/sites/un2.un.org/files/un-global-principles-for-information-integrity-en.pdf , p.2

[26] European Union External Action Service, Second EEAS Report on Foreign Information Manipulation and Interference (FIMI) threats, January 2024, p. 13, available at https://www.eeas.europa.eu/sites/default/files/documents/2024/EEAS-2nd-Report%20on%20FIMI%20Threats-January-2024_0.pdf

trustworthy, and reliable information, meaning that people can rely on the accuracy of the information they access while being exposed to a variety of ideas. [The] term "information integrity," [can] offer a positive vision of a broader information ecosystem that respects human rights and supports open, safe, secure, prosperous, and democratic societies" [27]. It also refers to the United Nations Policy Brief "Information Integrity on Digital Platforms", which states that "information integrity refers to the accuracy, consistency, and reliability of information. It is threatened by disinformation, misinformation and hate speech" [28]. While these definitions emphasize different aspects (the one provided by the Global Declaration focuses on a "positive vision" of the information ecosystem, and the one in the UN brief adopts a more "negative" perspective centered on the threats), they have one common feature: they are both concerned by the accuracy and reliability of the information that the people can access online. We believe this is not a useful or human rights-abiding framework for approaching phenomena like disinformation and election integrity. Truthfulness is not a relevant criterion to determine whether a piece of expression is protected under Human Rights Law. "Disinformation", "unreliable information," and "untrustworthy information" are labels that encompass both legal and illegal expression. Therefore, the new framing of old problems under the incipient "information integrity framework" does not provide any additional reason for state intervention and should be discarded.

16. We welcome that the Recommendation is clear in pointing out that the mere existence of risks "does not immediately or automatically lead to harm that requires regulation or the introduction of restrictive measures" [29] and in explaining that the adoption of such measures can pose a risk in themselves, especially for freedom of expression[30].
17. This is not to say that internet service providers and the activities they carry out must be immune from State scrutiny. On the contrary, as the Recommendation points out, the duty of the states to secure Human Rights to everyone within their jurisdiction under Article 1 of the European Convention demands them to take action against the violations that derive from their activity. States could be found in breach of these

[27] Government of the Netherlands (September 2023) Global Declaration on Information Integrity Online. Diplomatic Statement. https://www.government.nl/documents/diplomatic-statements/2023/09/20/globaldeclaration-on-information-integrity-online

[28] United Nations, Our Common Agenda, Policy Brief 8: Intormation Integrity on Digital Platforms, June 2023, available at https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrityen.pdf

[29] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 17.

[30] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 18.

obligations should they fail to take appropriate steps to prevent, investigate, punish, and redress abuse by private parties to persons under their jurisdiction[31]. We consider that the Recommendation provides positive insights insofar as they point to the unintended impacts of regulation, the three-part test, and the fact that disproportionate measures can themselves threaten Human Rights.

The limits of State action

18. We welcome that the Principles for Legal Frameworks contained in the Recommendation encourage legislators to adopt radically different approaches for dealing with the risks derived from lawful and unlawful content.
19. As for the procedures for the takedown of illegal content, we believe the Recommendation could be more emphatic in stating that restrictive measures must always follow a judicial determination regarding the illegality of the specific piece[32] and that adequate safeguards must be put in place to prevent disproportionate restrictions[33].
20. We commend that the Recommendation takes a holistic approach and encourages States to adopt content-neutral measures offline regarding the underlying societal conditions that cause online abuses[34]. Online-exclusive interventions centered on content (such as removal, deranking, demonetization, etc.) are usually ineffective, for they tackle a symptom rather than its causes, which often have deeper societal roots. We are pleased that it explicitly prohibits using "systemic" obligations *vis-à-vis* lawful content as ways to introduce content-specific restrictions that could not be otherwise lawfully imposed[35]
21. We appreciate that the Recommendation addresses the problem of incentives by warning against disproportionate sanctions and restrictions that could generate over-removal of legal content[36]. Co-regulatory regimes that lack proportionality guardrails, such as the Digital Services Act, put companies in the position of

[31] United Nations, Commentary to Principle 1, in Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, 2011

[32] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 39.

[33] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 46.

[34] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 20.

[35] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 28.

[36] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 45.

"voluntarily" removing much more legal content than they would otherwise remove if such incentives were not in place[37].

22. We are pleased that the Recommendation is clear in stating that any content rules that have an impact on Freedom of Expression should comply with the requirements of Article 10 of the European Convention on Human Rights[38]. We suggest further clarifying that also procedural, content-neutral rules should be subject to such a standard if they could be expected to have an impact on the way in which users receive or access information.
23. We also note with approval that the Recommendation acknowledges that states could be a source of risk to freedom of expression[39] by way of *jawboning* or other informal means[40].
24. While intermediaries must have legal resources at hand to defend themselves against state pressure[41], it is important that legal frameworks are designed in a way that minimizes the chances of it taking place. One of the ways the State can threaten freedom of expression is by participating in legally-mandated "voluntary" codes of conduct. State-led codes of conduct (such as those on Hate Speech and Disinformation under Article 45 of the DSA) deserve closer scrutiny, as they may allow regulators to impose restrictions on freedom of expression while circumventing human rights standards. These codes contain obligations that go way beyond what is required by International Human Rights Law, and therefore include actions that the states could not legitimately ask platforms to take[42]. While they are presented as legitimate efforts against dangerous expressions since platforms undertake such compromises formally voluntarily, their voluntary nature is disputed: from the outset, authorities pressured platforms to sign them under the threat of stricter regulations

[37] Agustina Del Campo, Nicolás Zara, and Ramiro Álvarez Ugarte, Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation, Research paper No. 64, Center for Studies on Freedom of Expression (CELE), Buenos Aires (2025). Forthcoming in the Journal of Intellectual Property, Information Technology and Electronic Commerce Law (vol. 16, No. 2), p. 21-22. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5161173.

[38] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 31, 35-38.

[39] Agustina Del Campo, Nicolás Zara and Ramiro Álvarez Ugarte, Reclaiming Human Rights for Platform Governance: Proposals for Restoring Their Centrality in the Era of Risks, Proceedings of Fourth European Workshop on Algorithmic Fairness, PMLR 294:265-280, 2025, p. 11.

[40] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 30.

[41] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 30.

[42] Natalie Alkiviadou, Why we should be concerned about the "Illegal hate speech" code of conduct +, blogpost at The Bedrock Principle, February 5, 2025, https://www.bedrockprinciple.com/p/why-we-should-be-concerned-about

and set the course their content should follow[43]. Moreover, there are both positive and negative incentives that make it too costly –almost inescapable– for platforms to stay out of them[44].

25. State-mandated trusted flaggers -like the ones mandated through Article 16 of the European Digital Services Act-, which the Recommendation encourages[45], are another opening for the threat of public overreach. The design of these mechanisms ought to exclude state actors or dependents from applying and acquiring such status. These systems must not be used by states to "*de facto* sidestep legal requirements to request content removal" or to intervene in content moderation without a legal basis, by flagging potential breaches of the platform's terms of service[46].

The role of users

26. We are pleased that the Recommendation refers to the individual responsibility of content creators to contribute to an online environment that is respectful of the rights of others. The definition of "content creators" in paragraph 11 is broad enough to reach almost any user of social media. However, even if the definition was intended to be narrower, we believe that they should get no special treatment based on their intended reach.
27. We agree that some users have heightened responsibilities based on their "particular role, professional standing, or position in society"[47] and that those who claim a certain expertise are bound by the ethical rules of their activity.
28. We believe that the best way to address many of the risks of harm generated online is by empowering users and reinforcing their autonomy. This includes, but is not limited to, allowing them to make more decisions on how their feeds should look like,

---

[43] Rachel Griffin and Carl Vander Maelen, "Codes of Conduct in the Digital Services Act: Exploring the Opportunities and Challenges" (May 30, 2023). Available at SSRN: https://ssrn.com/abstract=4463874 or http://dx.doi.org/10.2139/ssrn.4463874, also see Rachel Griffin and Carl Vander Maelen, Twitter's retreat from the Code of Practice on Disinformation raises a crucial question: are DSA codes of conduct really voluntary?, blogpost at the DSA Observatory Blog, June 12, 2023, https://dsa-observatory.eu/2023/06/12/twitters-retreat-from-the-code-of-practice-on-disinformation-raises-a-crucial-question-are-dsa-codes-of-conduct-really-voluntary/

[44] Íbid.

[45] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 82.

[46] Jacob van de Kerkhof, "Article 22 Digital Services Act: Building trust with trusted flaggers," Internet Policy Review: Journal on Internet Regulation, Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin, vol. 14(1), 2025, p. 8-9, retrieved from https://www.econstor.eu/bitstream/10419/315582/1/1923068466.pdf

[47] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 26.

permitting them to adjust their privacy settings, notifying and allowing for appeals when their content is taken down, etc[48].

Intermediary Liability and Due Diligence

29. States can and must pass legislation to govern the online ecosystem under their responsibility to protect and guarantee the enjoyment of human rights. But they must do so following strict human rights standards, which include freedom of expression and compliance with the three-part test: legality, necessity, and proportionality.

30. As a general rule, platforms must not be held liable for third-party-posted content, unless they do not comply with a judicial order for removal in a reasonable timeframe. The Recommendation should be explicit about the judicial nature of this notice in Paragraphs 42-44.

31. While we welcome the recommendation to impose due diligence obligations on platforms to assess whether their actions can have negative effects on human rights, we must warn against laws that conflate duties to identify and mitigate human rights risks with larger, overall social risks, dangerously broadening the language of the UNGPs and risking the existing consensus and agreements on human rights law worldwide. Some language in the recommendation, insofar as it requires platforms to assess impacts on "human rights, democracy and the rule of law"[49], could contribute to such confusion.

32. On the one hand, "risks to democracy and the rule of law" seems too broad a category to survive scrutiny under the principle of legality of Article 10 of the European Convention. On the other hand, invoking such a category may directly or indirectly affect freedom of expression—whether by targeting user-generated content, its distribution or recommendation, or the internal rules that platforms adopt for content governance. The only permissible basis for the state requiring platforms to remove, limit, deprioritize, or demonetize content is its illegality. Speech considered risky to democracy or the rule of law may still be entirely lawful. For this reason, the use of such a vague and expansive category must be rejected. At most, states could mandate platforms to act to protect important interests such as democracy and the rule of law from the risks stemming from legal expression, but they can do so only as long as those measures are content-neutral [50]. However, the mere exercise of the right to Freedom of Expression by the people should not be deemed "risky."

---

[48] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 55-56, 60-63.

[49] Council of Europe, Draft Recommendation CM/Rec(20XX)XX of the Committee of Ministers to member States on online safety and empowerment of content creators and users, par. 50.

[50] See Martin Husovec, The Digital Services Act's red line: what the Commission can and cannot do about disinformation, Journal of Media Law, 16(1), 47–56. https://doi.org/10.1080/17577632.2024.2362483.

Conclusions

33. We welcome the opportunity to participate in this consultation. The Recommendation has potential, yet could be further refined.

34. "Online safety" is a framework that does not provide clarity or certainty. Its incorporation into legal frameworks could have the undesired effect of narrowing down complex issues to private speech management by corporations. The definition of "online safety" provided by the Recommendation can and should be reformulated in the language of Human Rights.

35. The "information integrity" framework is equally inadequate and does not constitute a legitimate basis for state action.

36. The recommendations do a good job in warning against state overreach through risk-based approaches that incentivize overremoval of content. However, states could also threaten freedom of expression through other means, such as mandating companies to draft "voluntary" codes of conduct to manage legal expression and trying to obtain "trusted flagger" status.