Submission of the Citizen Lab to the Federal Government’s Proposed Approach to Address Harmful Content Online (“Online Harms Consultation”)

Cynthia Khoo, Lex Gill, and Christopher Parsons (Citizen Lab, Munk School of Global Affairs & Public Policy, University of Toronto)

Digital Citizen Initiative
Department of Canadian Heritage
25 Eddy St
Gatineau QC K1A 0S5

[*delivered electronically*]

25 September 2021

About the Citizen Lab and the Authors

1. The Citizen Lab at the Munk School of Global Affairs & Public Policy, University of Toronto (“Citizen Lab”), is an interdisciplinary laboratory which focuses on research, development, and high-level strategic policy and legal engagement at the intersection of information and communication technologies, human rights, and global security.

2. For over a decade, the Citizen Lab has used a mixed methods approach that combines techniques from network measurement, information security, law, and the social sciences to research and document information controls—including Internet censorship and surveillance—that impact the openness and security of digital communications and pose threats to human rights. Our work has investigated digital espionage against civil society; documented Internet filtering and other technologies and practices that impact freedom of expression online; analyzed privacy, security, and information controls of popular applications; and examined corporate and state accountability, transparency, oversight, and control in relation to information technologies, including the impact of those technologies on historically marginalized groups.

3. The Citizen Lab's groundbreaking research has resulted in over 120 publications and reports,[1] generated more than 25 front page exclusives in *The New York Times*, *Washington Post*, and other leading outlets, and received numerous international awards and recognitions. This scholarship has been cited by policymakers, academics, and civil society as foundational to the understanding of digital technologies, human rights, and global security.

4. The authors of this submission have diverse expertise in the freedom of expression, privacy, and/or equality rights implications of emerging technologies, including the specific kinds of technologies discussed in the "Technical Paper" associated with this consultation. We also have expertise in related areas of technology law and policy raised by this consultation, including adjacent questions of constitutional law, intermediary liability, privacy, national security, jurisdictional issues, and criminal evidence. We, alongside our colleagues at the Citizen Lab, have produced research on technologies closely related to this consultation, including consumer spyware apps ("stalkerware") and nation-state spyware, content filtering tools, anonymity tools, social media apps and platforms, and predictive policing and algorithmic surveillance technologies, among others. We have also published on issues related to corporate data collection, management, and disclosure and the relevant law and policy questions engaged by these issues. Each of us has routinely provided recommendations for technology, policy, and legal reform related to our respective findings in Canada and in various international fora.

5. We have reviewed the consultation materials, including the "Technical Paper" and the "Discussion Guide", associated with the government's proposal to address what it has referred to as "online harms".[2] We provide the following comments in response to that consultation process, divided into the following sections:

   A. This Consultation Is Inadequate;

   B. The Proposed Regime Will Not Achieve Its Intended Goals;

   C. The Scope of the Proposal Is Overbroad and Incoherent;

   D. Automated Enforcement Exacerbates Pre-Existing Problems;

   E. Unidirectional Takedown Incentive Will Likely Be Inequitable and Unconstitutional;

   F. Surveillance and Mandatory Reporting Requirements Are Dangerous and Chilling;

   G. New CSIS Powers Are Unjustified and Inappropriately Included in this Consultation; and

   H. Conclusion: Rewrite the Proposal from the Ground Up.

[1] A complete list of the Citizen Lab's publications, including research reports, articles, book chapters, resources and external submissions to government and international bodies is available online: <https://citizenlab.ca/publications/>.

[2] "Have your say: The Government's proposed approach to address harmful content online", Government of Canada (online): <https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content.html>.

A. This Consultation Is Inadequate

6. As a preliminary comment, the consultation process undertaken by the government on these proposals has been grossly inadequate.

7. The public materials, including the "Technical Paper", are vague, ambiguous, and in some cases contradictory. They fail to address some of the most obvious technical, legal, and constitutional problems they create—including problems that would negatively impact the purported "beneficiaries" of the proposed law. Some elements of the proposed measures are highly derivative of foreign legal regimes (including Germany's NetzDG regime and the United Kingdom's controversial Online Safety Bill), but lack the coherence or corresponding safeguards present in those schemes, and fail to respond to the criticism those regimes have faced on human rights grounds. Even the basic scope of what entities will be subject to the proposed measures is fundamentally uncertain because the definition of an "Online Communication Service Provider (OCSP)" is not sufficiently precise as presented,[3] and is subject to further change and indeterminacy through future regulations.

8. The materials also fail to offer any rational justification or evidence to demonstrate that the sweeping legal reforms proposed are likely to substantially mitigate the problems they purport to address. As a result, the materials lack a sufficient basis for comment and analysis, undermining the very function of a public consultation.

9. Though this consultation was preceded by a series of private, invite-only meetings between civil society groups and representatives from Heritage Canada and the Department of Justice, there is little to no evidence that the concerns raised by stakeholders at these meetings were accounted for in the government's proposal.[4] The government has been on notice that many of its proposals raise serious ethical, practical, and constitutional doubts since at least 2020, but these issues remain fundamentally unaddressed in the public materials. This failure to adjust course has undermined confidence among many experts and advocates that the function of the present consultation is, in fact, to consult, rather than to retroactively legitimize a series of foregone conclusions.

10. Finally, the period for written comments—particularly when limited to the end of summer during a federal election and global pandemic—has been insufficient to do justice to the sweeping proposals set out in the consultation materials. We would note that two of us are signatories to a public letter that

[3] For example, the "Discussion Paper" specifies that the category of OCSP is meant to "exclude travel review websites", yet such websites would seem to fit within the proposed definition of providing an OCS, a service that "enable[s] users of the service to communicate with other users of the service, over the internet", and the "Technical Paper" does not provide an explicit basis for exceptions from this definition.

[4] We would note that in December 2020, for example, two authors of this submission and a staff lawyer from another civil society organization met with Arif Virani (Parliamentary Secretary to the Minister of Justice and Attorney General of Canada) and Caroline Bourbonnière (Senior Advisor on Digital Policy to the Minister of Canadian Heritage) regarding this proposal. Many of the issues in this submission were flagged to the government's representatives at this time. One of the authors of this submission raised the same issues in speaking at a closed roundtable which also occurred in December 2020, where the Minister of Canadian Heritage Stephen Guilbeault was present, in addition to representatives of the same departments above.

protested the continuation of this consultation on the basis that it should not have proceeded after the government dissolved Parliament and called a federal election.[5] In our view, it was deeply inappropriate for this consultation to have continued during the caretaker period and we are disappointed that the government has failed to respond to these concerns.

B. The Proposed Regime Will Not Achieve Its Intended Goals

11. Technology-facilitated violence, abuse, and harassment is a real problem. Whether the violence, abuse, and harassment is based on gender (collectively, "TFGBV"), race, sexual orientation, other characteristics protected in Canadian equality law, or—more often than not—an intersecting combination of multiple characteristics, it plagues members of historically marginalized groups, who are routinely silenced and driven off the Internet as a result. This issue is serious and pressing, and it deserves and requires urgent and sustained attention from governments, technology companies, scholars, and civil society at every level.

12. In the same vein, thoughtless legislative measures to address these same issues for reasons of political expediency, or with insufficient care, thoughtfulness, intersectional and equitable considerations, and while lacking understanding of the practical and sociotechnical implications of such measures when implemented, do a profound disservice to the issue—as well as to targets, victims, and survivors, and to those historically marginalized groups whom online abuse, including NCDII and hate speech, most devastates.

13. In this respect, the proposals advanced by the government fail to account for the scholarship, concerns, and experiences of underrepresented, historically marginalized, and vulnerable individuals and communities. These are, of course, the very people who face the vast majority of technology-facilitated abuse, harassment, and violence—including women; Black, Indigenous, or otherwise racialized individuals; LGBTIQ+ individuals; individuals with disabilities; members of religious, linguistic and ethnic minority communities; immigrants and refugees; survivors of sexual violence, racist violence, and hate crimes; and sex workers—as well as individuals whose identities overlap multiple intersections among those groups.[6]

14. Research—including research produced by the Citizen Lab—has consistently demonstrated that Internet filtering and content monitoring technologies often result in the disproportionate censorship and surveillance of historically marginalized individuals and communities.[7] The technical interventions

[5] See OpenMedia et al, "Open letter: Defer consultations on the Internet until after the election" (2021), online: <https://openmedia.org/article/item/open-letter-requesting-rescheduling-of-open-internet-consultations>.

[6] See report and all sources cited within: Cynthia Khoo, "Deplatforming Misogyny: Report on Platform Liability for Technology-Facilitated Gender-Based Violence" (2021), at 23-28, online (pdf): *Women's Legal Education and Action Fund (LEAF)* <https://www.leaf.ca/wp-content/uploads/2021/04/Full-Report-Deplatforming-Misogyny.pdf>.

[7] See e.g., Jakub Dalek, Nica Dumlao, Miles Kenyon, Irene Poetranto, Adam Senft, Caroline Wesley, Arturo Filastò, Maria Xynou, and Amie Bishop. "No Access: LGBTIQ Website Censorship in Six Countries" (August 2021), Citizen Lab Research Report No. 142, University of Toronto, online: <https://citizenlab.ca/2021/08/no-access-lgbtiq-website-censorship-in-six-countries/>; Jakub Dalek, Lex Gill, Bill Marczak, Sarah McKune, Naser Noor, Joshua Oliver, Jon Penney, Adam Senft,

proposed by the Canadian government in the context of this consultation are emblematic of such an approach. The consultation materials advance an aggressive, algorithmic, and punitive regime for content removal it proposes, without any substantive equality considerations or clear safeguards against abuse of process. They also demonstrate the government's willingness to enlist and empower law enforcement and intelligence agencies to intervene on these issues—whether or not the victim or survivor has consented to such intervention.

15. In our view, any proposal that advances a superficial conception of safety for disadvantaged groups at the expense of their freedom to speak, create, relate, and organize, represents a false and potentially exploitative (and unconstitutional) promise. Furthermore, these measures encroach on individuals' right to privacy in serious ways, without substantially increasing "safety" in any case. This approach is predicated on a deeply paternalistic view that reduces vulnerable individuals to their right to security—again, a right that will not even necessarily be enjoyed under the proposed measures—rather than respecting the full constellation of their human rights and political entitlements, including the right to full and equal participation in democratic life.

16. The proposals similarly fail to account for the importance of protecting the kinds of expression that are most central to a free and democratic society—including journalism, academic scholarship and public interest research, debate, artistic creation, criticism, and political dissent, particularly when engaged in by members of historically marginalized groups. While the consultation purports to narrowly target five categories of already-illegal content,[8] there is almost no doubt that the proposed measures will have collateral consequences on lawful, democratic, and equality-advancing expression, including initiatives to document human rights violations,[9] creative forms of advocacy and protest, content that normalizes and celebrates the full diversity of sexual expression,[10] and efforts to de-escalate and counter violently extreme and harmful expression.

and Ron Deibert. "Planet Netsweeper" (April 2018), Citizen Lab Research Report No. 108, University of Toronto, online (pdf): <https://tspace.library.utoronto.ca/bitstream/1807/95393/1/Report%23108--Planet%20Netsweeper.pdf>; Ronald Deibert, Lex Gill, Tamir Israel, Chelsey Legge, Irene Poetranto, Amitpal Singh, "Submission to the UN Special Rapporteur on Violence Against Women, its Causes, and Consequences" (November 2017), online (pdf): *Citizen La*b <https://citizenlab.ca/wp-content/uploads/2017/11/Final-UNSRVAG-CitizenLab.pdf>.

[8] Though even this is not, strictly speaking, accurate, as the definitions proposed would likely encompass certain forms of content which is currently unambiguously lawful, and the definitions of key terms are subject to change through regulation.

[9] See e.g., Hadi Al Khatib & Dia Kayyali, "YouTube Is Erasing History", *New York Times* (23 October 2019), online: <https://www.nytimes.com/2019/10/23/opinion/syria-youtube-content-moderation.html>; and Belkis Wille, "'Video Unavailable': Social Media Platforms Remove Evidence of War Crimes" (September 2020), online: *Human Rights Watch* <https://www.hrw.org/report/2020/09/10/video-unavailable/social-media-platforms-remove-evidence-war-crimes>.

[10] See e.g., Cynthia Khoo, "Deplatforming Misogyny: Report on Platform Liability for Technology-Facilitated Gender-Based Violence" (2021), at 138-39, online (pdf): *Women's Legal Education and Action Fund (LEAF)* <https://www.leaf.ca/wp-content/uploads/2021/04/Full-Report-Deplatforming-Misogyny.pdf>.

C. The Scope of the Proposal Is Overbroad and Incoherent

17. The five categories of content identified by the government include "terrorist" content; content that incites violence; hate speech; non-consensual distribution of intimate images (NCDII); and child sexual exploitation content. These categories have little in common, beyond the fact they are illegal, and even then, the relevant legal analysis and basis for illegality is completely unique to each. In truth, the categories are united by almost nothing—constitutionally, factually, practically, or ethically—other than the proposed remedy of content removal.

18. In our view, any legislative scheme that purports to unite all of these disparate kinds of content under a single framework is incoherent, counterproductive, and constitutionally untenable. Each of these types of content implicates different *Charter* rights, operational considerations, risks of collateral harm from overbroad removal (as well as different risks of harm from under-removal), and different public policy concerns militating in favour of and against government intervention. The *Charter* and Canada's international human rights obligations require the government to engage in a proportionality analysis when restricting expression—weighing it against the nature and gravity of the harm that results, including the impact on other *Charter* rights such as the right to equality, as well as the government's legitimate interest in mitigating that harm. This analysis is, by design, extremely contextual.

19. Certain kinds of content addressed by the consultation present a strong constitutional foundation for expeditious removal powers, assuming appropriate safeguards are in place. NCDII is perhaps the clearest example, because the expressive value in such content is marginal and its categorization relies on an essentially straightforward analysis of whether or not the imagery was distributed with consent. But other forms of content require a much more nuanced evaluation. In particular, speech which appears to approach the statutory definitions of "terrorist" speech, incitement to violence, or hate propaganda is much more likely to intersect with legitimate acts of artistic expression, satire, irony, critique, parody, or in-group reappropriation of discriminatory words and imagery—all of which are entitled to constitutional protection.

20. Furthermore, it is essential to note that Canada's criminal law provisions regarding "terrorist speech" are constitutionally weak and largely untested.[11] For years, civil society organizations have raised the concern that these provisions—creatures of former Bill C-51, and later Bill C-59—have or could be used in a manner that exacerbates the wrongful criminalization and surveillance of Muslim and Arab individuals in particular.[12]

21. In our view, any future legislative scheme must be designed and justified in relation to *specific* harms and on the basis of *specific* government objectives rather than in relation to a generalized remedy. The

[11] For a more thorough discussion regarding the complexities of regulating "terrorist" content online and the constitutional vulnerabilities of the legislation currently in place, see Kent Roach, "Terrorist Speech under Bills C-51 and C-59 and the Othman Hamdan Case: The Continued Incoherence of Canada's Approach" (2019) 57:1 Alberta Law Review 203.

[12] See Part F ("Surveillance and Mandatory Reporting Requirements Are Dangerous and Chilling") below for elaboration on this point.

remedial powers associated with these schemes should then be vested within administrative tribunals and courts that have the capacity—or that are given the funding and resources to build capacity—and subject-matter expertise to properly weigh the issues at stake. Newly created "one-size-fits all" administrative bodies are an inappropriate forum to account for the complex and disparate concerns raised herein.[13]

D. Automated Enforcement Exacerbates Pre-Existing Problems

22. The consultation materials seem to encourage, if not all but mandate, the use of machine learning and similar automated technologies to enforce any legislated content regulations across OCSPs.[14] This drive towards automated enforcement may be rooted in the same rationale behind the government's decision to focus on the five types of content in question: a view that these categories are already relatively stable and narrowly circumscribed in Canadian law. To the extent that this is true in the law, the same cannot be said for these terms as they are likely to be interpreted or enforced by technology companies under the government's proposed approach—let alone through automated content moderation tools.

23. Before discussing the shortcomings of automated content moderation, it is worth establishing why human content moderation also poses difficulties. While courts, and to some degree expert administrative decision-makers, have generally proven capable of weighing context-sensitive and legal considerations on a case-by-case basis, the scale of the enforcement and removal envisioned by the consultation documents likely cannot, and will not, be achieved by human moderators.

24. Human reviewers of flagged content on social media are notoriously prone to error, due to factors such as problematic or misinterpreted company policies,[15] insufficient training, lack of time to properly assess content (as little as a matter of seconds), high-pressure environments that impose unimaginable stress, and lack of understanding of the content's cultural, social, or political context.[16] The reliance on human moderators across the digital platform industry is itself fraught; riddled with poor and

[13] That being said, one of the authors of this submission has advocated in other work the creation of a *specialized expert* administrative body that would focus solely on technology-facilitated violence, abuse, and harassment against members of historically marginalized groups, for similar reasons of these issues requiring sensitive and nuanced treatment given their complexity and the vulnerability of impacted individuals. We emphasize that this recommendation likewise explicitly rejects the "one-size-fits-all" approach proposed in the consultation materials. See Cynthia Khoo, "Deplatforming Misogyny: Report on Platform Liability for Technology-Facilitated Gender-Based Violence" (2021), at 225-27, online (pdf): *Women's Legal Education and Action Fund (LEAF)* <https://www.leaf.ca/wp-content/uploads/2021/04/Full-Report-Deplatforming-Misogyny.pdf>.

[14] "Technical Paper", at para 10.

[15] See e.g., Julia Angwin & Hannes Grassegger, "Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children", *ProPublica* (28 June 2017), online: <https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms>; and Samuel Gibbs, "Facebook bans women for posting 'men are scum' after harassment scandals", *Guardian* (5 December 2017), online: <https://www.theguardian.com/technology/2017/dec/05/facebook-bans-women-posting-men-are-scum-harassment-scandals-comedian-marcia-belsky-abuse>.

[16] See e.g., Kate Klonick, "Facebook Under Pressure", *Slate* (12 September 2016), online: <https://slate.com/technology/2016/09/facebook-erred-by-taking-down-the-napalm-girl-photo-what-happens-next.html>.

degrading working conditions; and characterized by low pay, professional insecurity, and psychological trauma.[17] Human content moderation on the scale required by the largest digital platforms raises serious issues of labour exploitation both in the home jurisdictions of major technology companies and internationally where such work is outsourced to third-party contractors in other countries.[18]

25. At the same time, automated content moderation such as through the deployment of machine learning algorithms fare little better. Examples abound of algorithmic errors in the content moderation field. They range from the superficially amusing, such as mistaking a photo of onions or of desert sand dunes for nudity,[19] to the politically disenfranchising—such as the algorithmic censorship of content by racial justice activists, adult content creators, sex education providers, documentors of war crimes and human rights violations, and political dissidents in authoritarian regimes.[20]

26. Further, algorithmic content moderation faces the same problems that inhere to nearly all forms of algorithmic decision-making, particularly in complex, contextual, socio-political environments. This includes the well-known issue of algorithmic bias—in particular, algorithmic bias against Black, Indigenous, and other racialized individuals and groups,[21] and gendered bias against women (both cis-

[17] See e.g., Sarah Emerson, "'A Permanent Nightmare': Pinterest Moderators Fight to Keep Horrifying Content Off the Platform", *OneZero* (28 July 2020), online: <https://onezero.medium.com/a-permanent-nightmare-pinterest-moderators-fight-to-keep-horrifying-content-off-the-platform-4d8e7ec822fe>.

[18] See generally Sarah T Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media* (New Haven and London: Yale University Press, 2019); and Elizabeth Dwoskin, Jeanne Whalen & Regine Cabato, "Content moderators at YouTube, Facebook and Twitter see the worst of the web—and suffer silently", *Washington Post* (25 July 2019), online: <https://www.washingtonpost.com/technology/2019/07/25/social-media-companies-are-outsourcing-their-dirty-work-philippines-generation-workers-is-paying-price/>.

[19] Coby Zucker, "Nudity algorithm wrongly blocked company's onion images, Facebook admits, says adverts will be restored", National Post (7 October 2020), online: <https://nationalpost.com/news/overtly-sexualized-st-johns-companys-onions-yes-onions-flagged-by-facebooks-nudity-algorithm>; and Melanie Ehrenkranz, "British Cops Want to Use AI to Spot Porn—But It Keeps Mistaking Desert Pics for Nudes", *Gizmodo* (18 December 2017), online: <https://gizmodo.com/british-cops-want-to-use-ai-to-spot-porn-but-it-keeps-m-1821384511>.

[20] See e.g., Danielle Blunt, Emily Coombes, Shanelle Mullin & Ariel Wolf, "Posting Into the Void" (2020), online: *Hacking//Hustling* <https://hackinghustling.org/wp-content/uploads/2020/09/Posting-Into-the-Void.pdf>; Shirin Ghaffary, "How TikTok's hate speech detection tool set off a debate about racial bias on the app", *Vox* (7 July 2021), online: <https://www.vox.com/recode/2021/7/7/22566017/tiktok-black-creators-ziggi-tyler-debate-about-black-lives-matter-racial-bias-social-media>; and Adam Smith, "Instagram Boss Says It Will Change Algorithm to Stop Mistreatment of Black Users, Alongside Other Updates", *Independent* (16 June 2020), online: <https://www.independent.co.uk/life-style/gadgets-and-tech/news/instagram-black-lives-matter-racism-harassment-bias-algorithm-a9567946.html>; Hadi Al Khatib & Dia Kayyali, "YouTube Is Erasing History", *New York Times* (23 October 2019), online: <https://www.nytimes.com/2019/10/23/opinion/syria-youtube-content-moderation.html>; Belkis Wille, "'Video Unavailable': Social Media Platforms Remove Evidence of War Crimes" (September 2020), online: *Human Rights Watch* <https://www.hrw.org/report/2020/09/10/video-unavailable/social-media-platforms-remove-evidence-war-crimes>; and Cynthia Khoo, "Deplatforming Misogyny: Report on Platform Liability for Technology-Facilitated Gender-Based Violence" (2021), at 138-39, online (pdf): *Women's Legal Education and Action Fund (LEAF)* <https://www.leaf.ca/wp-content/uploads/2021/04/Full-Report-Deplatforming-Misogyny.pdf>.

[21] The Citizen Lab has closely examined algorithmic bias in the context of algorithmic decision-making tools used to assess immigration and refugee applications, and to inform policing decisions and other parts of the criminal justice system. See Petra Molnar & Lex Gill, "Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System" (2018), online: *International Human Rights Program and the Citizen Lab* <https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf>; and Kate Robertson, Cynthia Khoo & Yolanda Song, "To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in

and trans-), non-binary individuals, and members of the LGBTIQ+ community.[22] Studies have shown that "hate speech detection" algorithms often demonstrate anti-Black racial bias,[23] and YouTube has gained a reputation for systematically hiding, demonetizing, or otherwise undermining LGBTIQ+ content on its platform.[24]

27. These difficulties are only exacerbated by the fact that people in Canada speak and access content in hundreds of different languages and distinct dialects, not all of which receive the same degree of resources or attention from technology platforms, but all of which would be subject to the government's proposed content removal regime.

E. Unidirectional Takedown Incentive Will Likely Be Inequitable and Unconstitutional

28. The incentive structure proposed by the government relies on large fines and other sanctions against technology companies to function. Combined with the 24-hour removal deadline, it is critical to note that the obligation and liability set out in the consultation materials appears largely unidirectional. The framework thereby rewards over-enforcement, with no countervailing forces to incentivize retention of content perceived as risqué or deviant by normative standards, but which remain legal, democratic, and often equality-advancing. This approach favours aggressive removal, the identification of false positives, and a risk-averse approach to sensitive or controversial content, as demonstrated by the empirical literature.[25] As emphasized above, it is the purported beneficiary groups of the proposed legislation—members of historically marginalized communities who are already silenced by both other users and the platforms themselves—who would disproportionately bear the brunt of wrongful takedowns. In our view, it is difficult to see how such an approach could be either equitable or constitutionally justifiable in Canada.

29. Moreover, 24 hours may be both too long *and* too short a window in which to require a platform company to act, depending on the type of content in question. The proposed legislation thus combines

Canada" (2020), online: *Citizen Lab and International Human Rights Program* <https://citizenlab.ca/wp-content/uploads/2020/09/To-Surveil-and-Predict.pdf>.

[22] Ari Ezra Waldman, "Disorderly Content" (17 August 2021) available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3906001>; and Daninel Leufer, "Computers are binary, people are not: how AI systems undermine LGBTQ identity" (6 April 2021), online: *Access Now* <https://www.accessnow.org/how-ai-systems-undermine-lgbtq-identity>.

[23] Maarten Sap et al, "The Risk of Racial Bias in Hate Speech Detection" in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (Florence: Association for Computational Linguistics, 2019) 1668 at 1668. See also Charlotte Jee, "Google's algorithm for detecting hate speech is racially biased" (13 August 2019), online: *MIT Technology Review* <https://www.technologyreview.com/2019/08/13/133757/googles-algorithm-for-detecting-hate-speech-looksracially-biased>.

[24] See e.g., Aja Romano, "A group of YouTubers is trying to prove the site systematically demonetizes queer content", *Vox* (10 October 2019), online: <https://www.vox.com/culture/2019/10/10/20893258/youtube-lgbtq-censorship-demonetization-nerd-city-algorithm-report>.

[25] See Daphne Keller, "Empirical Evidence of Over-Removal by Internet Companies Under Intermediary Liability Laws: An Updated List" (8 February 2021), online: *Center for Internet and Society (Stanford Law)* <https://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws>.

the worst of all worlds—potentially providing too little, too late in the case of NCDII, while courting unconstitutional overreach in the case of potential "hate speech" and "terrorist content". TFGBV experts consistently emphasize that speed is of the essence in the case of NCDII, given the danger in, devastating consequences of, and ease of downloading, reproducing, and further distributing the image or video, leading to further and repeated revictimization of the person depicted.[26] As mentioned above, identifying when something is NCDII (or child sexual exploitation) also poses fewer challenges compared to, in contrast, the likely more careful and nuanced analysis required for some situations of potential hate speech or "terrorist" content, for example.

30. Indeed, even Germany's NetzDG system, which has been deemed one of the more demanding platform regulation regimes, allows for up to seven days to assess and remove content that is not "manifestly unlawful".[27] Even to the extent that "hate speech" and "terrorist content" are unlawful, which has been the government's justification for their selection, it is far from the case that any given piece of content will manifestly fall within or outside of the relevant legal definitions. This is yet another instance demonstrating the incoherence, impracticality, constitutional fragility, and danger of addressing five legally, substantively, and sociopolitically different categories of content within the single blunt legal regime proposed. Addressing any of these issues in good faith requires separate, targeted legal regimes tailored to each category of content.

F. Surveillance and Mandatory Reporting Requirements Are Dangerous and Chilling

31. It is essential to understand that at a technical level, any requirement to proactively filter or proactively remove harmful content (i.e., in the absence of complaints) necessarily imposes obligations on platforms or internet service providers to engage in proactive monitoring of their users' content. In other words, platform liability regimes such as those proposed in the consultation materials may not only lead to corporate and proxy censorship, but may also amount to implementing platform-wide (or internet-wide) surveillance systems of user expression. The rights to privacy, equality, and freedom of expression are intertwined and thus interdependently threatened by the proposed regime in this respect.

32. In that light, the fact that the government's proposals would explicitly deputize technology companies in the surveillance and policing of their users on behalf of Canadian law enforcement and intelligence agencies is all the more disturbing. The proposed requirement on service providers to "take all reasonable measures, which can include the use of automated systems, to identify harmful content that is communicated on its OCS and that is accessible to persons in Canada, and to make that harmful content inaccessible to persons in Canada" appears to be nothing short of a positive legal obligation to

[26] See e.g., Nicola Henry & Asher Flynn, "Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support" (2019) 25:16 Violence Against Women 1932 at 1933; and Emily Laidlaw & Hilary Young, "Creating a Revenge Porn Tort for Canada" (2020) 96 Supreme Court Law Review 147 at 165.

[27] Heidi Tworek and Paddy Leerssen, "An Analysis of Germany's NetzDG Law" (15 April 2019) at 2, online (pdf): *Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression* <https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/06/NetzDG_TWG_Tworek_April_2019.pdf>.

monitor users and moderate their content. This approach will inevitably result in disproportionate levels of user censorship, including foreign users abroad with no relationship to Canada.[28]

33. Further, we are deeply concerned with the mandatory reporting requirements proposed in the consultation materials. The “Technical Paper” proposes to require that technology companies retain detailed records about their users, as well as report specific kinds of user activity directly to the RCMP and/or other law enforcement agencies. It also contemplates mandatory reporting of certain kinds of content to the Canadian Security Intelligence Service (CSIS). Specifically, the government proposes that “an OCSP shall report information respecting terrorist content and content that incites violence that will be made inaccessible in accordance with this legislation”.[29] Though this section of the consultation materials has been drafted in opaque language and provides broad, discretionary powers to Cabinet, it is clear that an automated mass informant scheme is what has ultimately been contemplated. These corporate informants, however, are essentially inescapable, given that they now play an almost infrastructural role in the social, relational, and political lives of people throughout Canada and around the world.

34. To this end, this proposal risks exacerbating the unconstitutional and discriminatory treatment of individuals whose information has been reported to law enforcement and/or CSIS. The reality is that these individuals are likely to belong to communities that are already disproportionately subjected to discriminatory over-criminalization by the police[30] and may be unjustly targeted for reporting to law enforcement. Such targeting and discriminatory treatment may be due to problematic or poorly applied platform policies, biased content moderation algorithms, or exploitation of the system by abusive users purposely targeting historically marginalized groups to drive them off a platform.[31]

35. We find it additionally troubling that how the government defines “terrorist content” varies throughout the consultation document, including the vague and recursive definition, “content that actively enourages terrorism and which is likely to result in terrorism”.[32] (Indeed, definitional issues are a recurring problem throughout the “Technical Paper” as a whole.) A loose or unclear definition of “terrorist content” raises particular issues regarding potential consequences on the rights of Indigenous peoples to express their views online and to organize protests or demonstrations, in the context of Indigneous land and water rights, Indigenous self-determination, the fraught issue of Canadian sovereignty, and the appropriation of Indigenous lands for resource extraction projects,

[28] The jurisdictional issues raised by this consultation are beyond the scope of our comments here, but we wish to acknowledge that they are both extensive and complex.

[29] “Technical Paper”, at para 22.

[30] See e.g., Kate Robertson, Cynthia Khoo & Yolanda Song, "To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada" (2020), at 15-18, online: *Citizen Lab and International Human Rights Program* <https://citizenlab.ca/wp-content/uploads/2020/09/To-Surveil-and-Predict.pdf>.

[31] See e.g., Suzie Dunn, "Technology-Facilitated Gender-Based Violence: An Overview" (2020), at 8, online (pdf): *Centre for International Governance Innovation* <https://www.cigionline.org/static/documents/documents/SaferInternet_Paper%20no%201_0.pdf>.

[32] “Technical Paper”, at para 8.

given similar concerns that many raised in response to earlier iterations of proposed national security legislation, Bill C-51 and Bill C-59.[33]

36. Perhaps most saliently, given the purported objectives of the proposed legislation, mandatory reporting requirements and automated involvement of law enforcement and intelligence agencies will, again, disproportionately harm and abrogate the fundamental rights and freedoms of historically marginalized groups. The proposed measures will result in widely chilling effects on such individuals' online activities and *dissuade* victims or survivors from seeking assistance if they believe that law enforcement may become involved—particularly when engaged without consent and outside of the impacted person's control.

37. At the heart of these concerns is the fact that the very groups who are systematically targeted for online abuse, and who are frequently the subjects of both actual and perceived hate speech, are the exact same groups who have been historically victimized or re-victimized and discriminated against by Canadian law enforcement and intelligence agencies.

38. For example, systemic discrimination and state violence against Black, Indigenous, and otherwise racialized people, by law enforcement and at all levels of the criminal justice system, has been thoroughly documented by impacted individuals, racial justice experts and advocates, human rights lawyers and researchers, the government itself at all levels, and multiple commissions, inquiries, and investigations over the course of decades.[34] On another front, national security and intelligence activities have been closely tied to Islamophobia and racial profiling against Arab and Muslim individuals, or those who are perceived to be Arab or Muslim, resulting in incursions on their ability to exercise constitutional rights and freedoms, including online.[35]

[33] See e.g., Doug Cuthand, "Bill C-51 has potential to scoop up aboriginal rights activists", *CBC* (6 May 2015), online: <https://www.cbc.ca/news/indigenous/bill-c-51-has-potential-to-scoop-up-aboriginal-rights-activists-1.3009664>; Hilary Beaumont, "The activists sabotaging railways in solidarity with Indigenous people", *Guardian* (29 July 2021), online: <https://www.theguardian.com/environment/2021/jul/29/activists-sabotaging-railways-indigenous-people>; Canadian Civil Liberties Association, "Submission to the Standing Committee on Public Safety and National Security regarding Bill C-59, *An Act respecting national security matters*" (January 18) at 14, online (pdf): *Canadian Civil Liberties Association* <https://ccla.org/wp-content/uploads/2021/06/2018-01-17-Written-submissions-to-SECU-re-C-59.pdf>; and International Civil Liberties Monitoring Group, "Brief on Bill C-59, the *National Security Act, 2017*" (May 2019) at 38, online (pdf): *International Civil Liberties Monitoring Group* <https://iclmg.ca/wp-content/uploads/2019/05/C-59-brief-May-2019-update.pdf>.

[34] See e.g., Kate Robertson, Cynthia Khoo & Yolanda Song, "To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada" (2020), at 15-28, online: *Citizen Lab and International Human Rights Program* <https://citizenlab.ca/wp-content/uploads/2020/09/To-Surveil-and-Predict.pdf>.

[35] See e.g., International Civil Liberties Monitoring Group, Islamic Social Services Association, and Noor Cultural Centre, "Islamophobia in Canada: Submission to the UN Special Rapporteur on Freedom of Religion or Belief" (November 2020), online: *OHCHR* <https://www.ohchr.org/Documents/Issues/Religion/Islamophobia-AntiMuslim/Civil%20Society%20or%20Individuals/Noor-ICLMG-ISSA.pdf>; Reem Bahdi, "No Exit: Racial Profiling and Canada 's War against Terrorism" (2003) 41:2-3 Osgoode Hall Law Jounrnal 293; Ashley Burke & Kristen Everson, "A Muslim former intelligence officer says systemic racism at CSIS is a threat to national security Social Sharing", *CBC* (29 June 2021), online: <https://www.cbc.ca/news/politics/racism-descrimination-claims-canadian-security-intelligence-service-1.6083353>; Tabasum Akseer, "Understanding the Impact of Surveillance and Security Measures on Muslim Men in Canada" (2018), at 45-85, online (pdf): *Centre for International and Defence Policy (Queen's University)*

39. The situation is such that victims/survivors of abuse, especially if they or the perpetrator are Black, Indigenous, or otherwise racialized or are vulnerable across multiple categories of oppression, will often avoid seeking aid from government institutions or calling the police because they do not want to be, or do not want the perpetrator to be, criminalized or subjected to police violence.[36] Tying automated police and national security agency intervention to their online spaces may only serve to isolate victims/survivors further, reducing their ability to seek help from their respective communities or through informal channels.

40. With respect to women (cis- and trans-), non-binary individuals, and other gender-diverse people, the disgraceful track record of law enforcement responses to both TFGBV and non-technology-facilitated sexual harassment and assault provides ample evidence to support fears of automated police and intelligence agency involvement in content moderation.[37] This is even more so considering that much online abuse and actual or perceived hate speech targeting women, gender-diverse people, and LGBTIQ+ individuals is sexualized, involves sexual harassment, or attempts to weaponize the targeted individual's sexuality against them.[38] Adding on a layer of technological and sociotechnical illiteracy among law enforcement,[39] in the context of online abuse and vulnerable marginalized individuals, portends nothing short of a recipe for disaster.

G. New CSIS Powers Are Unjustified and Inappropriately Included in this Consultation

41. For reasons that remain unclear, the government has seen fit to bury new powers for CSIS at the end of a paper ostensibly about platform regulation in relation to certain categories of harmful content,

<https://www.queensu.ca/cidp/sites/webpublish.queensu.ca.cidpwww/files/files/publications/Martellos/Martello42EN.pdf>; and Petra Molnar & Lex Gill, "Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System" (2018), at 19, online (pdf): *International Human Rights Program and the Citizen Lab* <https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf>.

[36] See e.g., Amanda Couture-Carron, Arshia U Zaidi & Nawal H Ammar, "Battered Immigrant Women and the Police: A Canadian Perspective" (2021) International Journal of Offender Therapy and Comparative Criminology 1; and Alexa Dodge, "Deleting Digital Harm: A Review of Nova Scotia's CyberScan Unit" (August 2021), at 22-23, online (pdf): *VAW Learning Network* <https://www.vawlearningnetwork.ca/docs/CyberScan-Report.pdf>.

[37] See e.g., Robyn Doolittle, "Unfounded: Why Police Dismiss 1 in 5 Sexual Assault Claims as Baseless", *Globe and Mail* (3 February 2017), online: <https://www.theglobeandmail.com/news/investigations/unfounded-sexual-assault-canada-main/article33891309>; Robyn Doolittle, "Unfounded: What It's Like to Report a Sexual Assault", *Globe and Mail* (17 March 2017), online: <https://www.theglobeandmail.com/news/investigations/what-its-like-to-report-a-sexual-assault-36-people-share-their-stories/article34338353>; and Cynthia Khoo, "Deplatforming Misogyny: Report on Platform Liability for Technology-Facilitated Gender-Based Violence" (2021), at 207, online (pdf): *Women's Legal Education and Action Fund (LEAF)* <https://www.leaf.ca/wp-content/uploads/2021/04/Full-Report-Deplatforming-Misogyny.pdf>.

[38] See e.g., Cynthia Khoo, "Deplatforming Misogyny: Report on Platform Liability for Technology-Facilitated Gender-Based Violence" (2021), at 16, online (pdf): *Women's Legal Education and Action Fund (LEAF)* <https://www.leaf.ca/wp-content/uploads/2021/04/Full-Report-Deplatforming-Misogyny.pdf>.

[39] See e.g., Cynthia Khoo, Kate Robertson & Ronald Deibert,, "Installing Fear: A Canadian Legal and Police Analysis of Using, Developing, and Selling Smartphone Spyware and Stalkerware Applications" (June 2019), at 165-67, online (pdf): *Citizen Lab* <https://citizenlab.ca/docs/stalkerware-legal.pdf>.

despite a tenuous relationship between these powers and the purported objectives of the consultation. There are several problems related to this portion of the proposal, which we discuss below.

42. First, in addition to the previously presented concern of violative and discriminatory intrusions into people's lives resulting from mandatory reporting, a related concern is the inability of impacted individuals to seek legal recourse where they are inappropriately targeted by CSIS through OCSPs. In the case of CSIS's security intelligence investigations (section 12 of the *CSIS Act*) or foreign intelligence investigations (section 16 of the *CSIS Act*), individuals who have their information disclosed to CSIS may never be able to contest such disclosures, or contest how that information is ultimately used by CSIS. This is in contrast to the context where law enforcement has brought criminal charges against an individual, and there is a greater possibility (comparatively speaking) of contesting the use of information which was disclosed to law enforcement based on information from an OCSP. Thus, the mandatory information sharing scheme proposed combines what will be almost certain inappropriate targeting of individuals by CSIS, with a negligible ability (as compared to in the law enforcement context) to seek legal recourse when unfairly impacted by such investigations. This concern might be mitigated if the government were to require OCSPs to transmit material exclusively to designated law enforcement agencies, instead of CSIS; however, we emphasize that there should be no mandatory reporting as described in the consultation materials in the first place, and that this was an entirely inappropriate context in which to seek expanded powers for CSIS.

43. Second, Canadian academics have robustly demonstrated that Canada suffers from a severe "intelligence to evidence" problem that is often linked to CSIS being unable or unwilling to communicate information to law enforcement bodies due to concerns that doing so will compromise sources or methods.[40] This results in either defendants in criminal cases being robbed of their due process rights, or the inhibition of criminal prosecutions where there is otherwise reason for them to proceed. Again, this issue might be addressed by limiting OCSP information-sharing to law enforcement agencies, but we stress that any legislation purporting to set up new information-sharing channels among or between law enforcement, intelligence agencies, and digital platforms or other technology companies *must* be the subject of its own dedicated public consultation process.

44. Third, we oppose the proposal to grant CSIS a new warranting power on the basis of the consultation materials provided. If the *CSIS Act* were modified, as proposed, the Service would broaden its foreign intelligence operations collection capacity by being able to collect basic subscriber information without having to satisfy section 21 warranting requirements, compared to if section 21 were reformed

[40] For some of this discussion, see: Kent Roach, "The Unique Challenges of Terrorism Prosecutions: Towards a Workable Relation between Intelligence and Evidence" in *Commission of Inquiry into the Investigation of the Bombing of Air India Flight 182* (2010), available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1629227>; Kent Roach, "The eroding distinction between intelligence and evidence in terrorism investigations" in *Counter-Terrorism and Beyond*, eds Nicola McGarrity, Andrew Lynch & George Williams (Abington: Routlege, 2010); Leah West, "The Problem of 'Relevance': Intelligence to Evidence Lessons from UK Terrorism Prosecutions" (2018) 41:4 Manitoba Law Journal 57; Craig Forcese & Leah West, "Threading the Needle: Structural Reform & Canada's Intelligence-to-Evidence Dilemma" (2019) 42:4 Manitoba Law Journal 131; and Dave Murray & Derek Huzulak, "Improving Intelligence to Evidence (I2E) Model in Canada" (2021) 44:1 Manitoba Law Journal 181.

to include a data production power. This is extremely problematic, firstly, on the grounds that the online harms consultation should not seek to reform how foreign intelligence operations are undertaken and, secondly, on grounds that the government has not demonstrated a clear reason for why data production powers cannot be added to existing section 21 restrictions, as opposed to being provided with a reduction in the court's oversight concerning such new proposed powers. We also note that expanded powers were already granted to CSIS not even five years ago, in the passage of Bill C-59.[41]

45. It is not evident from the "Technical Paper" why basic subscriber data production powers could not be added to the existing section 21 regime, as opposed to under a separate unique regime. While the government indicates an issue with timeliness in conducting investigations and a desire for flexibility in operations, it has not demonstrated that section 21 is actually impeding investigations. Again, should this be the case, then the government should hold formal public consultations on national security as opposed to integrating these debates within a broader consultation that primarily concerns completely separate areas of law, scholarship, and expertise. The totality of online expression collectively captured by the consultation materials' proposed five categories far exceeds the national security context or the scope of CSIS's mandate.

46. There may be a debate worth having about CSIS's ability to obtain basic subscriber information pertaining to foreign actors that are allegedly engaged in terrorism-related activities associated with inciting terrorist attacks. However, this consultation—which is, at its heart, focused on the role and responsibilities of digital platforms in the context of intermediary liability law, content moderation, online abuse, and technology-facilitated gender-based violence, abuse, and harassment—does not provide the appropriate forum to do so.

47. Fourth, we oppose the proposed *CSIS Act* modification on grounds that the proposed reforms would weaken Federal Court oversight of CSIS operations at a time where the Service has exhibited a chronic failure to meet its existent obligations to behave with candour towards the courts.[42] There is ample public evidence demonstrating that CSIS and its counsel have actively misled the Federal Court in relation to CSIS's activities and operations. This misconduct should not be rewarded with weakened judicial oversight to obtain new classes of information by way of evading the requirements present under the existing section 21 regime. At the very least, any data production powers that are meant to facilitate section 16 activities should fall under section 21 of the *CSIS Act* instead of operating under a separate regime—but again, no national security legislative reform should occur as a result of this consultation.

[41] Catharine Tunney, "Canada's national security landscape will get a major overhaul this summer", CBC (23 June 2019), online: <https://www.cbc.ca/news/politics/bill-c59-national-security-passed-1.5182948>.

[42] See e.g., Jim Bronskill, "Court admonishes CSIS once again over duty of candour", *Globe and Mail* (31 August 2021), online: <https://www.theglobeandmail.com/canada/article-court-admonishes-csis-once-again-over-duty-of-candour>.

H. Conclusion: Rewrite the Proposal from the Ground Up

48. For all of the reasons detailed above, we strongly recommend that the federal government revise, in earnest, this particular piece of proposed legislation, from the ground up. It is not too late to change course, and to incorporate recommendations that reflect what civil society groups and technology and human rights experts have been communicating directly to the responsible ministries and departments over the course of the past several years—alongside representatives of the purported beneficiaries of the proposed legislation, such as historically marginalized groups targeted by TFGBV.

49. At the very least, the proposed measures should be broken up into two or more separate pieces of legislation—if not a separate legal regime tailored to each of the five designated categories, then perhaps one specific to NCDII and/or child sexual exploitation materials, and a separate scheme (or schemes) addressing hate speech, terrorist content, and/or incitement to violence. That separation would result in more internally coherent proposals, fewer constitutional complications, and more honest and precise debates regarding each of the five content categories on their own merits, rather than being dangerously and counterproductively conflated with each other. The government would be more likely to achieve its purported objectives in that case, whereas the currently proposed measures will only set up all involved for failure, at the expense of those already being harmed the most.