



LEAF
FAEJ

WOMEN'S LEGAL
EDUCATION & ACTION FUND
FONDS D'ACTION ET D'ÉDUCATION
JURIDIQUE POUR LES FEMMES

Submission to Canadian Heritage on the Federal Government's Proposed Approaches to Address Harmful Content Online

25 September 2021

Written by: Moira Aikenhead, Suzie Dunn, and Rosel Kim¹

¹ Moira Aikenhead is a PhD Candidate at the Peter A. Allard School of Law at the University of British Columbia. Suzie Dunn is an Assistant Professor at Dalhousie University's Schulich School of Law and is a Senior Fellow with the Centre for International Governance Innovation. Rosel Kim is a Staff Lawyer at the Women's Legal Education and Action Fund. Jane Bailey, Cynthia Khoo, Ngozi Okidegbe, and Karen Segal also contributed to this submission.

Table of Contents

Introduction	3
About LEAF and its Expertise	3
Issues with the Regulatory Framework from a Substantive Equality Perspective	5
A. Lack of Substantive Equality Framework.....	6
B. Lack of Consultation.....	8
C. Overbreadth and Inextricable Connection with the Criminal Justice Process.....	9
D. The Timeframe for Takedown Requirements Must be Tailored for Different Harms.	11
E. Necessity of Transparency and Disaggregated Data from Platforms on All Instances of TFGBV.....	12
F. Issues Regarding a Digital Safety Commissioner	13
G. Increasing the Focus on Education and Prevention	17
H. Dangers of Algorithmic Moderation.....	18
Appendix: Full List of Recommendations from “Deplatforming Misogyny”	19

Introduction

The Women’s Legal Education and Action Fund (LEAF) supports the development of a federal regulatory framework to address the growing issue of technology-facilitated gender based violence (TFGBV), which disproportionately impacts historically marginalized communities, including women, girls, and gender-diverse people. However, we do not support the federal government’s proposed “online harms” framework as drafted, because it poses serious concerns from a substantive equality and human rights perspective and risks exacerbating existing inequalities, particularly because it purports to deal with five very different “online harms” with a single approach.

LEAF believes that in order to deal effectively with the growing issue of TFGBV, the government must allocate resources to create a regulatory framework dealing exclusively with it as a particular harm. **We urge the government to: a) revise the regulatory framework to explicitly recognize substantive equality and human rights as guiding principles; b) to provide more immediate and direct support to victims experiencing TFGBV; c) to provide alternative remedies to those provided through law enforcement and the criminal justice process; d) to recognize forms of TFGBV that are not currently captured by the criminal law; and e) to ensure responses are tailored to and account for the specific harms of TFGBV.**

About LEAF and its Expertise

LEAF is a national, charitable, non-profit organization that works towards advancing substantive gender equality through litigation, law reform, and public education. Since 1985, LEAF has intervened in over 100 cases—many of them before the Supreme Court of Canada—that have advanced equality rights in Canada.

Some forms of online harms—as defined by the proposed framework—directly engage LEAF’s mandate of substantive gender equality. Conduct such as hate speech and non-consensual distribution of intimate images (NCDII) have a disproportionately detrimental impact on women and gender-diverse people’s ability to express themselves and participate without fear in many online spaces that have become crucial to our professional and personal lives. These harms are the ones we will address in this submission.

LEAF has developed expertise in the gendered impact of online hate and TFGBV. In 2019, LEAF intervened in the landmark case of [R v Jarvis](#),² where it urged the Supreme Court of

² 2019 SCC 10.

Canada to apply an equality lens when interpreting the *Criminal Code* provision of voyeurism. LEAF has also made submissions to Parliament to highlight the gender equality implications of hate speech and online hate, such as its submission to the House of Commons Standing Committee on Justice and Human Rights' study of online hate in 2019.³ In April 2021, LEAF released a research report "[Deplatforming Misogyny](#)"⁴ by human rights and technology lawyer Cynthia Khoo, which examines how digital platforms can be held accountable and liable for their role in perpetuating TFGBV from a substantive equality perspective.

In "Deplatforming Misogyny", LEAF made 14 recommendations for federal action to regulate TFGBV, including legislative reform. These recommendations are based on 6 guiding priorities that emerged from the research and analysis conducted in this report and should govern efforts to address TFGBV in Canadian law. These priorities are:

1. recognizing a need for legal reform to address TFGBV, including through platform regulation;
2. recognizing that Canadian constitutional law justifies imposing proportionate limits on freedom of expression in order to uphold and protect the rights to equality and freedom from discrimination, and also to give full effect to the core values underlying freedom of expression;
3. guaranteeing that legal reforms that address TFGBV build in victim/survivor-centered, trauma-informed, and intersectional feminist perspectives;
4. ensuring expedient, practical, and accessible remedies for those targeted by TFGBV;
5. providing due process mechanisms to users who wish to contest platforms' content moderation decisions (whether a decision to leave up or take down content); and
6. requiring transparency from platform companies regarding their content moderation policies and decisions, as well as the outcomes of such policies and decisions concerning TFGBV.

³ Women's Legal Education and Action Fund, "Submission to the House of Commons Standing Committee on Justice and Human Rights Respecting the Committee's Study of Online Hate" (2021), online (pdf): *Women's Legal Education and Action Fund* <<https://www.leaf.ca/wp-content/uploads/2019/05/2019-05-10-LEAF-Submission-to-the-Standing-Committee-on-Justice-and-Huma....pdf>>.

⁴ Cynthia Khoo, "Deplatforming Misogyny" (April 2021), online (pdf): *Women's Legal Education and Action Fund* <<https://www.leaf.ca/publication/deplatforming-misogyny/>>. [*Deplatforming*]

Of LEAF's fourteen recommendations for federal action, we emphasize the following six that are relevant to our submission (see the [Appendix](#) of this submission or the "[Deplatforming Misogyny](#)" report for full list):

1. Establish a centralized expert regulator **for TFGBV specifically**, with a dual mandate: a) to **provide legal remedies and support to individuals** impacted by TFGBV on digital platforms, including regulatory and enforcement powers; and b) to develop research on TFGBV and **provide training and education to the public, relevant stakeholders, and professionals**.
2. **Ensure that legislation addressing TFGBV integrates substantive equality considerations** and guards against exploitation by members of dominant social groups to silence expression by members of historically marginalized groups.
3. Ensure that **legislation to address TFGBV focuses solely on TFGBV (including intersectional considerations)—do not dilute, compromise, or jeopardize the constitutionality of such legislation by 'bundling' TFGBV with other issues** that the government may wish to also address through platform regulation.
4. Enact a law that allows for victims/survivors of TFGBV to obtain **immediate removal of certain clearly defined kinds of content from a platform without a court order**, such as the non-consensual distribution of intimate images.
5. Require platform companies to undergo independent audits (which could be conducted by the new TFGBV agency) and **publish comprehensive annual transparency reports**.
6. **Fund frontline support workers and community-based organizations** working to end, and supporting victims/survivors of, gender-based violence, abuse, and harassment, **specifically to enhance their internal expertise, resources, and capacity to support those impacted by TFGBV** (which often accompanies gender-based violence and abuse).

Issues with the Regulatory Framework from a Substantive Equality Perspective

It is LEAF's view that regulating hateful, discriminatory, and harmful content is necessary and important for enhancing freedom of expression and equality rights. LEAF also believes that the government must play a central role in regulating TFGBV, because technology companies that operate digital platforms have not demonstrated willingness to safeguard the rights of free expression for all users. In fact, recent investigations into business decisions of digital platforms demonstrate how the companies are ignoring the evidence of harm that users

experience, because online hate and harassment can be particularly lucrative.⁵ It is therefore inadequate to leave the work of regulating profitable, yet harmful, online content, including many forms of TFGBV, to the very companies who stand to gain from that content. These realities underscore the broader need for regulation of industry practices that themselves incentivize and perpetrate online hate, harassment and discrimination. The government has a crucial role to play in prioritizing Canadians' constitutional rights over corporate growth. In order for the regulatory framework that addresses TFGBV to be effective, **it must be grounded in substantive equality and human rights.**

For the reasons cited above, we encourage a governmental regulatory framework—including an expert regulatory body—which we believe is necessary for protecting the freedom of expression and equality rights of all, especially for women and gender-diverse people. However, the proposed regulatory framework is inadequate and raises several concerns from an equality perspective. Our comments will focus on the types of online harms included in the framework that LEAF has expertise in: hate speech and NCDII. We will also include some comments on child sexual abuse material (CSAM).

As explained below, **we find it highly problematic that such distinct harms as NCDII would be addressed under the same legislation as other offences such as terrorism, and believe each of the harms the framework proposes to address require unique approaches and considerations.**

We outline our concerns in detail below:

A. Lack of Substantive Equality Framework

Any regulation of hateful or harmful speech **must adopt an explicitly intersectional and substantive equality lens.**⁶ It must recognize that discriminatory, threatening and other harmful speech targets and silences marginalized voices.⁷ Research in Canada and abroad show that when women and other marginalized groups are faced with discriminatory and hateful speech, non-consensually distributed images, and attacks when speaking out about

⁵ Georgia Wells, Jeff Horwitz and Deepa Seetharaman, “Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show”, *Wall Street Journal* (14 September 2021), online: <https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=article_inline>.

⁶ *Deplatforming*, *supra* note 4 at 222-223.

⁷ See *Saskatchewan (Human Rights Commission) v Whatcott*, 2013 SCC 11 at para 114.

equality issues, one of the main consequences is for the targeted groups to engage less or to stop engaging online.⁸

Adopting substantive equality principles requires acknowledging the ways regulatory frameworks and content moderation processes can be abused by dominant groups to further silence marginalized voices.⁹ Beyond silencing marginalized voices and forcing members of these groups offline, hateful online rhetoric has resulted in significant tangible harms and violence to these groups. Hateful and discriminatory online speech has been connected to some of Canada's deadliest attacks including the Toronto van attack and the Quebec City mosque shooting where the attackers had a history of following sexist online groups (incels) who promote violence against women and hateful Islamophobic online groups, respectively, prior to their attacks and posting sexist and racist content online.¹⁰ The focus of regulating TFGBV in digital spaces must centre squarely on the abuse of historically marginalized groups, advancing their equality and upholding their human rights.¹¹

The government must commit to addressing hateful speech and violence targeting these equality-seeking groups by making this intent explicit in the regulatory framework, and the substance of the framework should reflect that intention. The framework proposed in the Technical Paper acknowledges that online hate has disproportionate impacts on marginalized groups including women, Indigenous Peoples, members of racialized and religious minority communities and LGBTQ2 and gender-diverse communities and persons with disabilities. It is ostensibly premised on respecting and protecting the ability of people to fully participate in the public discourse free from harm.¹²

The regulatory framework requires Online Content Service Providers (OCSPs) to ensure that the implementation of measures to make harmful content inaccessible does not

⁸ See Amnesty International, "Toxic Twitter - A Toxic Place for Women" (March 2018) online: *Amnesty International* <<https://www.amnesty.org/en/latest/news/2018/03/online-violence-against-women-chapter-1/>>; Anastasia Powell & Nicola Henry, *Sexual Violence in a Digital Age* (London: Palgrave Macmillan UK, 2017); Plan International, "Free to be Online? A report on girls' and young women's experiences of online harassment" (October 2020) online: *Plan International* <<https://plan-international.org/publications/freetobeonline>>.

⁹ *Deplatforming*, *supra* note 4 at 224.

¹⁰ Stephane J Baele, Lewys Brace & Travis G. Coan, "From 'Incel' to 'Saint': Analyzing the violent worldview behind the 2018 Toronto attack" (2019) *Terrorism and Political Violence*, DOI: [10.1080/09546553.2019.1638256](https://doi.org/10.1080/09546553.2019.1638256); Michael Nesbitt, "Violent crime, hate speech, or terrorism? How Canada views and prosecutes far-right extremism (2001-2019)" (2021) 50:1 *Common L World Rev* 38 <<https://journals.sagepub.com/doi/full/10.1177/1473779521991557>>.

¹¹ *Deplatforming*, *supra* note 4 at 225.

¹² Government of Canada, "Technical Paper" (29 July 2021) at 1.c.; 1.h, online: *Canadian Heritage* <<https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content/technical-paper.html>>. [*Technical Paper*]

“result in differential treatment of any group on a prohibited ground of discrimination.”¹³ Once again, the government must explicitly acknowledge and name what groups are likely to be targeted and adversely impacted by TFGBV, and guarantee that any measures to identify and render inaccessible harmful speech must not operate to silence voices that have been historically marginalized. Research has shown that women and gender-diverse people, particularly those with intersecting marginalities such as race, sexual orientation, gender expression and disability face increased online attacks based on their identity and often respond to these attacks by engaging less online.¹⁴

B. Lack of Consultation

Despite some communication with organizations and individuals with expertise on TFGBV, the Government failed to consult meaningfully with experts, civil society groups, victims and survivors of TFGBV in crafting the proposed framework. To the extent that the government held consultations and conversations with a select group of organizations and individuals (as it did with LEAF), the current proposal does not reflect many of the recommendations or concerns that were raised. For example, LEAF raised concerns about the extensive information-sharing powers that were proposed during a joint call with Feminist Alliance for International Action (FAFIA) and Heritage staffers on February 10, 2021. It also raised concerns about the focus on criminal law enforcement and encouraged the government to take a substantive equality approach to this legislation.

In order to ensure any proposed legislation does not result in unintended adverse consequences for equality-seeking groups, **the government must meaningfully consult with groups and organizations with specialization in these areas before tabling any proposed legislation governing the regulatory framework, Digital Safety Commissioner, or Digital Recourse Council of Canada and Advisory Board.** We understand the Liberal Government intends to introduce legislation governing online harms within its first 100 days.¹⁵ We strongly caution against such an approach as a significantly more robust and meaningful consultation with impacted groups is required, which is not possible in the proposed 100-day timeline.

¹³ *Ibid* at 10.a.

¹⁴ See Amnesty International, “Toxic Twitter - A Toxic Place for Women” (March 2018) online: *Amnesty International* <<https://www.amnesty.org/en/latest/news/2018/03/online-violence-against-women-chapter-1/>>.

¹⁵ Liberal Party of Canada, “Forward. For Everyone: Protecting Canadians from Online Harms.” (2021), online: *Liberal* <<https://liberal.ca/our-platform/protecting-canadians-from-online-harms/>>.

C. Overbreadth and Inextricable Connection with the Criminal Justice Process

Like other forms of gender-based violence, TFGBV is rooted in intersecting and systemic oppressions including misogyny, racism, colonialism, homophobia, transphobia and ableism.¹⁶ **It is crucial that legislation aimed at addressing TFGBV explicitly focus on TFGBV specifically, and not conflate it with other forms of “online harms”.** While there is an urgent need to address other forms of what the framework has defined as “online harms”, it is inappropriate to conflate these distinct issues under one single approach. “Bundling” regulation of this form of TFGBV with other types of content moderation, such as speech related to terrorism, compromises the utility and integrity of such a framework, and jeopardizes its constitutionality.¹⁷

By proposing to regulate five separate and distinct categories of harmful content (child sexual exploitation; terrorist content; content inciting violence; hate speech; and the non-consensual distribution of intimate images), each of which requires its own unique response, the proposed framework is overbroad and prevents the specified tailoring needed to adequately respond to each of these unique issues. Each of the “online harms” identified in the framework must be addressed individually, particularly when it comes to reporting mandates, types of support available for victims, and the necessity and timing of content removal.

The five categories of harmful content covered by the framework are each subject to the provisions of the *Criminal Code*.¹⁸ While the definition of “hate speech” is to be informed by the definition under the *Canadian Human Rights Act*, the hate speech provisions under that legislation are not yet in force, and the proposed definition closely mirrors the interpretation of “hate propaganda” pursuant to the *Criminal Code*.¹⁹ While the framework suggests definitions must be adapted to the regulatory context, it is not clear what this means.²⁰ **This framework should develop definitions of TFGBV that are grounded in substantive equality, rather than the criminal law,** which has not always centered gender equality.²¹ The

¹⁶ *Deplatforming*, *supra* note 4 at 225.

¹⁷ *Ibid* at 228.

¹⁸ *Criminal Code*, RSC 1985, c C-46.

¹⁹ Bill C-36, *An Act to amend the Criminal Code and the Canadian Human Rights Act and to make related amendments to another Act (hate propaganda, hate crimes and hate speech)*, 2nd Sess, 43rd Parl, 2020-2021 (first reading 23 June 2021).

²⁰ *Technical Paper*, *supra* note 12 at 8.

²¹ See e.g.: Emma Cunliffe, “Sexual Assault Cases in the Supreme Court of Canada: Losing Sight of Substantive Equality?” (2012) 57 SCLR (2d) 295 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2111652>; Margaret Denike, “Sexual Violence and ‘Fundamental Justice’: On the Failure of Equality Reforms to Criminal Proceedings” (2000) 20:3 Canadian Woman Studies 151 <<https://cws.journals.yorku.ca/index.php/cws/article/viewFile/12681/11764>>.

framework's definitions must emphasize the need to regulate content that interferes with the equality rights of those targeted by the content. For example, the framework should target forms of TFGBV targeting women, girls, and gender-diverse people.

The framework represents a missed opportunity to provide support to victim/survivors of TFGBV by failing to address some of the most common forms of problematic content online that does not meet the definitions set out in the *Criminal Code*, but still causes significant harm. For example, rape threats and death threats are frequently aimed at equality-seeking groups, however, many of these threats may not reach the criminal definition of uttering threats. It should also be noted that many forms of TFGBV that proliferate online were not included in the list of “online harms”, such as harassment and threats. It is not clear why the five specific harms were selected for regulation and not others. Further, criminal definitions of hate speech and NCDII cast a fairly narrow net, excluding many kinds of hateful online commentary and exploitative images that do not fit in the definitions within the *Code*. Content that does not rise to the level of criminality can still cause serious harm to marginalized groups.²² The government must provide mechanisms to support groups targeted by content that does not meet the definition of criminality but is nevertheless harmful, such as providing support in navigating social media's content moderation procedures and providing emotional and technical support, particularly if that content breaches a platform company's own content moderation rules.

Aligning the definition of “online harms” with *Criminal Code* offences, and requiring mandatory reporting to law enforcement in certain circumstances, does not align with a victim/survivor-centric, intersectional, or substantive equality approach to regulating TFGBV. Victims/survivors should have some element of choice when seeking support. For some, this may involve a criminal justice response, while others will be better served through less formal support, such as a government help line or victim service support workers providing technical and emotional support. Members of marginalized communities that have a history of being over-criminalized or having their complaints ignored or neglected by the police. Particularly, Black, Indigenous, and racialized communities, and women reporting sexual violence, may be reluctant to engage in a regulatory system that requires reporting to the police due to factors such as over-criminalization, even if the content they are concerned with is criminal in nature.

In cases involving TFGBV such as NCDII, **LEAF recommends that there can be no mandatory reporting to law enforcement without the express informed consent of the**

²² The United Kingdom, Secretary of State for Digital, Culture, Media & Sport and the Secretary of State for the Home Department, “Online Harms White Paper” (April 2019), online (pdf): *GOV.UK* <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/973939/Online_Harms_White_Paper_V2.pdf>.

victim/survivor.²³ Mandatory reporting may be appropriate and required under existing legislation for cases involving child sexual abuse material. For adult women, a mandatory reporting regime may deter some from seeking help because not everyone wishes to engage the police. While police involvement may be necessary in many cases, the framework should take into consideration groups of individuals who may not seek help at all because of concerns about, fear of, or prior negative experiences with, police involvement - especially for those who are Black, Indigenous, and racialized. The failure to build in consent from an individual victim or target illustrates the shortcomings of this regulatory framework, which attempts to regulate disparate harms such as terrorism and incitement of violence with non-consensual distribution of intimate images.

Mandatory reporting risks the over-criminalization of individuals and puts innocent people at risk of being reported to the police. Further, as noted in Alexa Dodge's report on CyberScan, mandatory reporting to the police has not proven to be effective in addressing certain forms of online harms, such as NCDII among young people, many of which will involve victim/survivors who do not want police involvement.²⁴ These harms must be taken seriously in all cases, but providing trauma-informed, survivor-centered options, rather than mandating police involvement in all cases, is essential to providing effective remedies for survivors.

While not strictly within LEAF's mandate, we note the serious risks to the substantive equality of marginalized groups in potentially requiring the mandatory flagging and reporting of terrorist content or content that incites violence, as each risks capturing content created and promoted by marginalized groups protesting state violence, and over-policing racialized communities, particularly if algorithms are used to identify content and mandatory reporting policies are in place. The requirement that OCSOs ensure their notifications to law enforcement do not result in differential treatment on a prohibited ground²⁵ is insufficient.

D. The Timeframe for Takedown Requirements Must Be Tailored for Different Harms

Expedient removal of harmful online content must be balanced with freedom of expression interests and aim to avoid over-removal and wrongful takedowns. For this reason, each of the five categories of "online harms" in the framework require different timelines. For sexual images shared without consent, timeliness of removal is of utmost importance for those featured in the images.²⁶ Unlike other forms of online harms identified within the framework,

²³ *Deplatforming*, *supra* note 4 at 227.

²⁴ Alexa Dodge, "Deleting Digital Harm: A Review of Nova Scotia's CyberScan Unit" (August 2021), online (pdf): *VAW Learning Network* <<http://www.vawlearningnetwork.ca/docs/CyberScan-Report.pdf>> [*Digital Harm*].

²⁵ *Technical Paper*, *supra* note 11 at 8.

²⁶ Emily Laidlaw & Hilary Young, "Creating a Revenge Porn Tort for Canada" (2020) 96 SCLR (2nd) 147.

intimate images and child sexual abuse materials are much easier to identify in content removal procedures. The harms caused by their distribution are vastly increased the longer the content stays online and is available to be downloaded, viewed and shared by others. The salutary effects of swift content removal outweigh the deleterious effects to the free expression of those affected by those immediate takedowns. This balancing will not be the same for the other harms listed in the proposal. **We support expedient takedown requirements for child sexual abuse material and NCDII and images, have profound impacts on the equality interests of those targeted in these images, the effects of which are amplified as the images are proliferated.** Further, these images are easier to identify than other forms of harmful online content discussed in the framework, reducing the likelihood of over-removal or wrongful takedown of such images.

However, any proposed legislation, and those responsible for implementing and administering it, must remain alive to how an expedient takedown rule for NCDII could negatively impact sex workers and other people who are expressing themselves sexually online. The regulatory body must invest resources in the relevant expertise to distinguish abusive and exploitative distribution of intimate images from instances where groups or individuals are misusing complaint mechanisms to attack and silence those with non-normative sexual identities, or those who engage in consensual non-normative sexual practices.²⁷ “Deplatforming Misogyny” provides an example of such negative impacts when discussing the consequences of US Senate bill *Stop Enabling Sex Traffickers Act* (SESTA) and the House Bill *Allow States and Victims to Fight Online Sex Trafficking Act* (FOSTA), enacted in April 2018, which resulted in social media companies prohibiting and removing vast amounts of legitimate sexual expression content - including sex education materials - in order to protect themselves from liability under these statutes which were intended to prohibit exploitative content.²⁸

E. Necessity of Transparency and Disaggregated Data from Platforms on All Instances of TFGBV

We also support imposing an obligation for OCSPs to provide regularly scheduled reports to the Digital Safety Commissioner about their content moderation practices.²⁹ However, **we urge the government to expand the OCSPs’ reporting obligations by requiring them to submit disaggregated demographic data in all instances of TFGBV (such**

²⁷ See Ari Waldman, "Disorderly Content" (16 August 2021), online (pdf): Available at SSRN <<https://ssrn.com/abstract=3906001> or <http://dx.doi.org/10.2139/ssrn.3906001>> [*Disorderly Content*].

²⁸ *Deplatforming*, *supra* note 4 at 139.

²⁹ *Technical Paper*, *supra* note 12 at para 14.

as NCDII and hate speech) so that researchers and civil society organizations can accurately glean how misogyny, racism, ableism, homophobia, and other forces of oppression are impacting the platforms. Currently, the framework only requires OCSPs to provide disaggregated data when the incidents of online harm are shared with law enforcement.³⁰ Content moderation transparency reporting obligations should not be tied to reports to law enforcement agencies, nor, as noted above, should OCSPs be required to report every incident of TFGBV to law enforcement.

One of the recommendations in “Deplatforming Misogyny” was to require digital platform companies to “undergo independent audits [...] and publish comprehensive annual transparency reports.”³¹ We also recommended that the transparency reports that the data in the report “should be broken down by demographics (particularly gender and race) to the extent possible, regarding the platform’s internal content moderation policies and practices, and regarding the prevalence of and efforts to address TFGBV, as well as the results of those efforts.”³² Requiring data from platforms will be critical for academics and civil society organizations to understand OCSP’s content moderation practices and gather the relevant information in order to conduct research and/or advocate for equality-centered reform.

Regulatory requirements including content moderation reporting needs to differentiate between platforms of various sizes, natures, purposes and business models. Regulations should not be so burdensome that it will prevent smaller sites or companies from complying with them or beginning at all. This consideration needs to be taken into account on all aspects that the government seeks to legislate.

F. Issues Regarding a Digital Safety Commissioner

Taking into consideration the critiques made above, **we support the establishment of a government body or bodies such as the Digital Safety Commissioner and Digital Recourse Council of Canada, so long as the government revises the approach of the regulators to an equality-based one.** This means that the regulator must have the objective of providing accessible and meaningful remedies to those targeted by TFGBV and actively seeking to adjust norms and behaviours around TFGBV through public education and evidence-based research.

The current framework has some positive aspects to it, including:

³⁰ *Ibid* at para 14 (h)(II).

³¹ *Deplatforming*, *supra* note 4 at p. 229 (Recommendation #10).

³² *Ibid*.

- Requirements for the Commissioner to engage with groups disproportionately affected by harmful online content;
- Requirements that Online Communication Service Providers (OCSP) provide reports of their content moderation practices;
- Requirements that social media companies have clear content moderation guidelines;
- Inclusion of a formal complaints process for individuals to make complaints of non-compliance with regulations and failure to follow content moderation guidelines;
- Administrative monetary penalties (AMP) for ongoing non-compliance;
- Recognizing that “hatred spread online often has a disproportionate impact on women, Indigenous Peoples, members of racialized and religious minority communities and on LGBTQ2 and gender-diverse communities and persons with disabilities” and that “that OCSs are used to sexually exploit children online, and that such exploitation can have life-long consequences for victims”;
- Requirements that flagged content be addressed expeditiously (however, the current approach on timing must be reexamined depending on the content flagged);
- Supports for platforms in reducing harmful content;
- Engagement in partnerships, education outreach activities, and research on TFGBV;
- Requirements that members of the commission, council and advisory body have subject matter expertise and are inclusive of marginalized communities and groups protected under the *Canadian Human Rights Act*.

The framework for these bodies could be improved by taking the following considerations into account:

- **The proposed regulatory bodies should provide accessible and immediate supports for victim/survivors of TFGBV as well as more systemic responses.**
- Governmental bodies that provide targets of TFGBV with direct support, administrative support, and educational campaigns have proven to be useful to those individuals impacted by TFGBV.³³ **Research has shown what is most commonly needed by victim/survivors of TFGBV is immediate technical safety support such as support in getting content taken down as well as emotional support and information from people with subject matter expertise.**³⁴
- In many cases of non-consensual distribution of intimate images, what victim/survivors need is immediate support navigating social media company’s content moderation

³³ Pam Hrick, *The Potential of Centralized and Statutorily Empowered Bodies to Advance a Survivor-Centered Approach to Technology-Facilitated Violence Against Women* (Bingley, UK: Emerald Publishing, 2021) [*Centralized and Statutorily Empowered Bodies*].

³⁴ *Digital Harm*, *supra* note 24.

processes and other tactics for getting content removed.³⁵ Even with statutory regulations that require timely content removal, victims/survivors will need accessible information and direct assistance in reporting and understanding reporting procedures and will need supports beyond simply getting the content taken down.

- The federal government should look to bodies such as Nova Scotia’s CyberScan, New Zealand’s Netsafe, Australia’s eSafety Commissioner and the UK Revenge Porn Helpline as examples of government supported initiatives that provide immediate help to targets of TFGBV.³⁶ These bodies provide help lines, direct reporting mechanisms, and information that provide immediate support to those targeted by TFGBV and other forms of problematic behaviour online.
- These bodies have staff who understand social media companies’ reporting systems and can provide assistance in getting content removed. Non-consensually distributed intimate images and child sexual abuse material is already prohibited by most major social media sites. When these organizations have established relationships with the major social media companies where the bulk of harms occur, they can provide more direct support than an individual can. For example, in 2018, the eSafety Commissioner of Australia was able to have 90% of the NCDII reported to them removed.³⁷
- The proposal requires that prohibited content not be available in Canada. This is unclear whether the content will be deleted, as is necessary for NCDII and CSAM, rather than blocked for Canadian users through geolocation/IP filtering. NCDII and CSAM must be deleted and not be accessible to any users.
- These supports should be buttressed with regulatory requirements that social media companies remove particularly harmful forms of content in a timely manner and penalties for failing to do so. This is necessary because many social media companies will otherwise be incentivized to allow harmful forms of TFGBV to remain on their platforms if this content is lucrative or drives up user engagement.³⁸
- Statutory and regulatory requirements, along with government supported bodies that can provide immediate support and remove the burden from targets and place greater obligations on platforms that have to date failed to adequately address TFGBV on an individual and systemic level.

³⁵ *Ibid.*

³⁶ *Centralized and Statutorily Empowered Bodies*, *supra* note 33.

³⁷ Australian Government, “Annual Reports 2018-19: Australian Communications and Media Authority Office of the eSafety Commissioner” (2018-19), online: *ACMA* <https://www.esafety.gov.au/sites/default/files/2019-10/ACMA_and_eSafety_annual_reports_2018_19.pdf>.

³⁸ *Deplatforming*, *supra* note 4 at 53-54; See also Georgia Wells, Jeff Horwitz and Deepa Seetharaman, “Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show”, *Wall Street Journal* (14 September 2021), online: <https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=article_inline>.

- These organizations should provide technical and emotional support to those harmed by content that is not illegal, such as getting content removed that breaches a social media company's content moderation policies or providing emotional and technical support to those who have been targeted by TFGBV.
- These direct support mechanisms should be accessible 24 hours a day and should provide phone, email and texting options so the harms can be addressed at the time that they occur.
- It should be noted that non-normative and LGBTQ+ content is more likely to be inappropriately flagged, taken down and banned. As such, there also need to be timely mechanisms in place to challenge when this content is inappropriately flagged and made inaccessible in order for that content to be made accessible again³⁹ through an accessible and timely counter notification process.⁴⁰
- Requiring victims/survivors to engage in a regulatory process that will take weeks or months is not a viable solution for people whose sexual images have been posted online. As noted by Alexa Dodge, when there are few supports and a complex system to report TFGBV, people are unlikely to engage in the very systems meant to protect them.⁴¹
- In more extreme cases where the perpetrator refuses to take down content, social media companies fail to properly implement their content moderation guidelines, or content is posted on websites that are dedicated to hosting TFGBV, such as revenge websites, additional government support is needed. In these situations, a formalized process through a digital safety commissioner or other body could be helpful to address these issues.
- In cases where the victim/survivor is interested in pursuing a criminal response, meaningful support for them should be available. It is well documented that victims of sexual violence have experienced unsupportive and discriminatory responses from some criminal justice system actors, including police officers and the courts. Any regulatory body set up to address TFGBV should work with those in the criminal justice system to ensure they do not mistreat or revictimize women and others who have been the targets of TFGBV. In all cases, the choice of whether or not to engage with the criminal justice system must remain with the victim/survivor, and not forced upon them.

³⁹ *Disorderly Content*, *supra* note 27.

⁴⁰ Sonja Solomun, Maryna Polataiko & Helen Hayes, "Platform Responsibility and Regulation in Canada: Considerations on Transparency, Legislative Charity, and Design" (2021) 34 Harv JL & Tech <<https://jolt.law.harvard.edu/digest/platform-responsibility-and-regulation-in-canada-considerations-on-transparency-legislative-clarity-and-design>>.

⁴¹ *Digital Harm*, *supra* note 24.

- Victim/survivors should be empowered to choose their own course of action and should have multiple courses of action, including formal and informal responses.⁴² This may include flagging and removing content, engaging with law enforcement agencies when the behaviour is criminal, and/or speaking with a specialist in TFGBV who can provide emotional and technical support to manage the incident.
- The educational material produced by this body should encourage a cultural shift in attitudes toward TFGBV specifically and gender-based violence. This should be done in the school system and for the larger public.⁴³
- Ongoing research should be conducted on TFGBV to understand trends and examine the effectiveness of government responses, including regulatory and criminal ones. This research must be informed by evidence based on the experiences of victim/survivors.⁴⁴

G. Increasing the Focus on Education and Prevention

An effective regulator of online harm must not only provide remedies to online harm that occurs, but also proactively seek to change the culture by educating the public and the decision-makers about the oppressive roots of TFGBV.⁴⁵ For this reason, we urge the government to identify research and education as one of the central mandates of the Commissioner to prevent future acts of TFGBV.

Currently, education only gets a cursory mention in the Technical Paper, which states the Digital Safety Commissioner will engage in “[p]artnerships, education and outreach activities, and research” to help fulfill the policy objectives of the new legislation.⁴⁶

The research and education function should not be just geared towards government and academics, but to the public at large. In order to effectively serve a preventative function, the education materials should be publicly accessible in format and content.

⁴² *Centralized and Statutorily Empowered Bodies*, *supra* note 33.

⁴³ *Ibid.*

⁴⁴ *Ibid.*

⁴⁵ Anastasia Powell et al., “Image-based sexual abuse: An international study of victims and perpetrators. A Summary Report” (February 2020) at 12, online (pdf): *RMIT University* <https://researchmgt.monash.edu/ws/portalfiles/portal/319918063/ImageBasedSexualAbuseReport_170220_WEB_2.pdf>.

⁴⁶ *Technical Paper*, *supra* note 12 at para 35(b).

H. Dangers of Algorithmic Moderation

As anticipated in the framework, algorithmic identification of harmful content will be required by larger platforms to comply with the framework's requirements. This raises serious equality concerns, as **algorithms have not proven to be failsafe mechanisms for identifying and removing harmful content, particularly content that requires detailed analysis such as hate speech**. This adds a significant risk of over-compliance and the removal of legitimate content.

Proactive algorithmic based removal of child sexual abuse material may be appropriate in most circumstances, however, in cases of NCDII algorithms will not be suited to identify the difference between legitimate images of sexual expression and those posted without consent. In the case of NCDII, content that is flagged as NCDII should be removed immediately.

Additionally, the use of algorithmic moderation poses serious substantive equality concerns. There is significant scholarship about the discriminatory impact of these systems.⁴⁷ In the realm of content moderation, the distributive harms stemming from the use of these algorithms will largely be experienced by gender-diverse, racialized, Indigenous and other marginalized communities.⁴⁸ This has already occurred in the United States in the context of hate speech; research from 2019 showed that AI models for detecting hate speech online were more likely to flag tweets from Black posters as offensive or hateful.⁴⁹ The reason for this is that algorithms do not produce neutral outcomes; instead, the outcomes reflect the biases of their designers and the biases contained in the data that they are constructed and trained on.⁵⁰

We would welcome an opportunity to discuss any of the above further.

⁴⁷ For discussing the role of bias in data and what can be done about it, see: Solon Barocas & Andrew D. Selbst, "Big Data's Disparate Impact" (2016) 104 Cal L Rev 671

<https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899>

⁴⁸ Robert Gorwa, Reuben Binns & Christian Katzenbach, "Algorithmic content moderation: Technical and political challenges in the automation of platform governance" (28 February 2020), online: *Big Data & Society* <<https://journals.sagepub.com/doi/full/10.1177/2053951719897945>>.

⁴⁹ See for example Shirin Ghaffary, "The algorithms that detect hate speech online are biased against black people" (15 August 2019) online: *Vox* <<https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter>>

⁵⁰ Sandra G Mayson, "Bias in, Bias out" (2019) 128 Yale LJ 2218 <https://www.yalelawjournal.org/pdf/Mayson_p5g2tz2m.pdf>.

Appendix: Full List of Recommendations from “Deplatforming Misogyny”

From Cynthia Khoo, “Deplatforming Misogyny” (April 2021), online (pdf): *Women’s Legal Education and Action Fund* <<https://www.leaf.ca/publication/deplatforming-misogyny/>>:

Guiding Priorities and Recommendations for Federal Action

This report provides 14 recommendations for federal action, including legislative reform. These recommendations are based on six guiding priorities that emerged from the research and analysis conducted in this report and should govern efforts to address TFGBV in Canadian law.

These priorities are:

- recognizing a need for legal reform to address TFGBV, including through platform regulation;
- recognizing that Canadian constitutional law justifies imposing proportionate limits on freedom of expression in order to uphold and protect the rights to equality and freedom from discrimination, and also to give full effect to the core values underlying freedom of expression;
- guaranteeing that legal reforms that address TFGBV build in victim/survivor-centered, trauma-informed, and intersectional feminist perspectives;
- ensuring expedient, practical, and accessible remedies for those targeted by TFGBV;
- providing due process mechanisms to users who wish to contest platforms’ content moderation decisions (whether a decision to leave up or take down content); and
- requiring transparency from platform companies regarding their content moderation policies and decisions, as well as the outcomes of such policies and decisions concerning TFGBV.

Recommendations for Federal Action

A. Centering Human Rights, Substantive Equality, and Intersectionality

1. Apply a principled human rights-based approach to platform regulation and platform liability, including giving full effect to the rights to equality and freedom from discrimination.
2. Ensure that legislation addressing TFGBV integrates substantive equality considerations and guards against exploitation by members of dominant social groups to silence expression by members of historically marginalized groups.

3. When pursuing legislative or other means of addressing TFGBV, consult substantively with and take into account the perspectives and lived experience of victims, survivors, and those broadly impacted by TFGBV.

B. Legislative Reforms

4. Establish a centralized expert regulator for TFGBV specifically, with a dual mandate: a) to provide legal remedies and support to individuals impacted by TFGBV on digital platforms, including regulatory and enforcement powers; and b) to develop research on TFGBV and provide training and education to the public, relevant stakeholders, and professionals.
5. Enact one or more versions of the current ‘enabler’ provision in subsections 27(2.3) and 27(2.4) of the Copyright Act, adapted to specifically address different forms of TFGBV, including ‘purpose-built’ platforms.
6. Enact a law that allows for victims/survivors of TFGBV to obtain immediate removal of certain clearly defined kinds of content from a platform without a court order, such as NCDII.
7. Ensure that legislation to address TFGBV focuses solely on TFGBV (including intersectional considerations)—do not dilute, compromise, or jeopardize the constitutionality of such legislation by ‘bundling’ TFGBV with other issues that the government may wish to also address through platform regulation.

C. Legal Obligations for Platform Companies

8. Require platform companies to provide to users and non-users clearly visible, easily accessible, plain-language complaint and abuse reporting mechanisms to expediently address and remedy instances of TFGBV.
9. For ‘purpose-built’, ‘enabling’, or otherwise TFGBV-dedicated platforms, and where a clearly delineated threshold of harm is met, provide that an order to remove specific content on one platform will automatically apply to any of that platform’s parent, subsidiary, or sibling platform companies where the same content also appears.
10. Require platform companies to undergo independent audits (which could be conducted by the new TFGBV agency) and publish comprehensive annual transparency reports.
11. When determining legal obligations for digital platforms, account for the fact that platforms vary dramatically in size, nature, purpose, business model (including non-profit), extent of intermediary role, and user base.

D. Research, Education, and Training

12. Fund, make widely available, and mandate (where appropriate) education resources and training programs in TFGBV, which include information on how to support those who are subjected to TFGBV.
13. Fund frontline support workers and community-based organizations working to end, and supporting victims/survivors of, gender-based violence, abuse, and harassment, specifically to enhance their internal expertise, resources, and capacity to support those impacted by TFGBV (which often accompanies gender-based violence and abuse).
14. Fund further empirical, interdisciplinary, and law and policy research by TFGBV scholars, other TFGBV experts, and community-based organizations on TFGBV and the impacts of emerging technologies on those subjected to TFGBV.