

The Government's Proposed Approach to Address Harmful Content Online

Submission to the Government of Canada

September 2021

Founded in 1951, the Canadian Association of University Teachers (CAUT) is the national voice for academic staff representing 72,000 teachers, librarians, researchers, general staff, and other academic professionals at some 125 universities and colleges across the country. CAUT is an outspoken defender of academic freedom and works actively in the public interest to improve the quality and accessibility of post-secondary education in Canada.

As defenders of academic freedom, the right to teach, research, publish and express opinions without fear of political or institutional censorship, CAUT has grave concerns about the online harms bill the government intends to introduce in the autumn of 2021.

At this time, CAUT urges the government to reconsider this legislative project. The proposed approach described in the consultation's technical paper and discussion guide is rife with unintended, serious, and harmful consequences. The kind of regime being considered by the government would inadvertently censor legal speech and undermine the rights and civil liberties of Canadians.

Proposed approach

The proposed framework requires content platforms—Online Communication Service Providers (OCSP)—to police and remove content that falls into one of the five categories of “online harms.” This includes the use of machine learning and Artificial Intelligence (AI) to proactively search for harmful content. Users could also indiscriminately flag any content as potentially being illegal. Once content is flagged, OCSPs will have 24 hours to remove it. OCSPs will also have to report content they remove directly to the RCMP, CSIS or both without notifying the user.

A new ‘Digital Safety Commissioner’ would be created to oversee this regime, though they would not report to Parliament. This is problematic because, among other things, the Commissioner would have the power to conduct hearings on content takedowns in secret, justified by privacy, commercial and industrial secrecy, national security and defense, and international relationships with other governments.

The harmful content targeted by this legislation is wide-ranging and poorly defined. It includes terrorist content; content that incites violence; hate speech; non-consensual sharing of intimate images; and child sexual exploitation content. CAUT acknowledges the deep harm this content can cause, especially to vulnerable individuals and marginalized groups. We do question whether the proposed approach is indeed the best tool at this time to address the problems of these online harms given that they are already offences in Canada's *Criminal Code*. CAUT is also doubtful whether the introduction of one single regulatory regime is the best way to address the variety of online harms targeted by this legislation.

Systemically flawed

The imposition on OCSPs to determine what is lawful content is problematic. The proposed approach is systemically flawed to incentivize OCSPs to be over-vigilant and over-remove content. Some ways the government has designed this system to encourage hyper-vigilance on the part of OCSPs include:

- The speed with which OCSPs would be required to remove flagged content (24 hours);
- The sheer volume of content that would have to be moderated; and,
- Stiff penalties which the Digital Safety Commissioner would be empowered to impose (whichever is greater of either 3% of a OCSP's global revenues or \$10 million dollars.)

Further indications that the government's proposal is systemically tilted towards censorship includes the proposal for a ‘Digital Recourse Council’. This body of 3-5 people would hear appeals from users regarding OCSP's moderation decisions. The Council's decisions, curiously, would be binding in the instance of OCSP content takedowns but non-binding for the re-instatement of content.

Another dimension for consideration in the proposed approach is the potential for it to be used by malicious internet actors as a tool to silence and abuse innocent individuals and communities, particularly those who are already marginalized. Giving users the opportunity to report on others can be weaponized, especially by

organized groups of internet vigilantes or crusaders operating to advance a particular viewpoint or political agenda. Though user reporting is already a mainstay of OSCP moderation today, responsible OSCPs cannot necessarily do their due diligence in responding to these reports if the government imposes upon them the added pressure of speed and financial penalties. The result is that OSCPs will be incentivized to remove content and lock accounts of innocent parties under attack from internet trolls.

Distinguishing between legal & illegal content

Distinguishing between legal speech and illegal content is not always simple and obvious. In our democratic society, much that is awful, is likely also lawful speech, as the courts have set a high bar for what constitutes prohibited hate speech. Nonetheless, the ability to distinguish between lawful and illegal is difficult; even the courts struggle to do this with legal experts, rigorous arguments, and an ample amount of time for open and transparent inquiry.

The difficulties of making this distinction between illegal content and legal speech are only exacerbated when the task is given to machine learning and AI, which cannot necessarily understand the entire context in which content exists and operates.

- Examples of how legal content might be misidentified and removed by OSCPs if algorithms fail to fully grasp the context of content and statements.
- Academic researchers investigating unpopular or controversial topics may use OSCPs to exchange and share information. This new legislation and onus on OSCPs to police and remove content could have an impact on academic research and extra-mural speech.
- Protest literature, sociopolitical satire, conflict photography¹, or the documentation of human rights abuses could undermine civil disobedience

and censor voices looking to bring important nuance and debate to sensitive subject matters.

- Artists, museums, galleries, and art educators use image content, like nude art, to promote exhibits, public lectures, and other research that could be misidentified as sexual content.² Quick content takedowns and the lengthy complaint and recourse regime could have a significant impact on the ability to promote events that are substantial revenue generators for those working in the cultural sector.
- Vulnerable individuals and marginalized groups frequently come together in online spaces to find community, seek out support and discuss their experiences. If these discussions include relaying information about experiences of discrimination or attacks, AI surveillance could wrongfully flag this content as online harm.

In the last example, censoring this legal speech would have the unintended consequence of exacerbating the existing, well-documented pattern of online speech policing and removal targeting equity-deserving individuals and communities. Further to this point, relying on machine learning and AI could perpetuate social inequities given issues around algorithmic biases and insufficient access to the full breadth of training data used by OSCPs.

Privacy concerns & unwarranted surveillance

Moderating and decontextualizing online content is further complicated when considering that OSCPs, under the proposed approach, are required to report to security agencies when content is flagged harmful, opening the door to unwarranted surveillance of academics and researchers. Whether through human or AI-generated moderation, under this scheme the government is incentivising private companies to moderate, make determinations of, and share data and

1. The iconic Pulitzer Prize winning photo of the naked Vietnamese nine-year-old girl running away from a napalm attack was misidentified by Facebook as child pornography and taken down in 2016. It took an international backlash for the platform to reverse

its decision. See, BBC News "[Fury over Facebook 'Napalm girl' censorship](#)" (09 September 2016).

2. Hyperallergic "[Facebook Censors Art Historian for Posting Nude Art, Then Boots Him from Platform](#)" (27 November 2018).

information on suspected criminal activity, without alerting affected individuals.

The proposed legislative and regulatory framework would only further institutionalize and grant security agencies with powers to collect data and monitor information about Canadians, with no commensurate increase in oversight or accountability. Academics and researchers could be subjected to surveillance creating a chill on political discourse that challenge dominant paradigms. The technical paper provides little clarity on limitations to interagency information sharing or time limits for how long security agencies are permitted to collect and store data and information.

Summary

CAUT supports net neutrality, the principle that Internet Service Providers should enable access to all content and applications regardless of the source, and without favoring or blocking particular products or websites. The development of an open Internet has been instrumental in dramatically expanding both research capability and learning opportunities for Canadian academics, researchers, and students. The government's proposed approach to addressing harmful content online has serious shortcomings regarding protecting principles of net neutrality and open internet. The threat of website blocking, proposed as a punitive measure for OCSP deemed noncompliant, is a direct violation of net neutrality.

Other problematic areas identified in the government's proposed regime include national security accountability and oversight, and risks to the open exchange of information and infringement of basic civil liberties. It's worth noting that net neutrality and the Charter of Rights and Freedoms are never mentioned once in the technical paper and discussion guide. The concerns highlighted in this submission need to be more fulsomely discussed with stakeholders and better nuanced to protect rights and freedoms while addressing legitimate concerns over online criminal activity.

CAUT strongly recommends more extensive consultation, including rescheduling roundtable discussions, to find a way forward to protect against discrimination, harassment, and violence, while avoiding regulating expression that may offend some, but is lawful.