

CCNC-SJ Submissions to the Virtual Roundtable with Minister Steven Guilbeault on proposed regulations for social media platforms on countering online harm

December 8, 2020

The Chinese Canadian National Council for Social Justice (“CCNC-SJ”) has been invited by the Heritage Minister Steven Guilbeault to make submissions on the proposed regulation to regulate social media platforms as they relate to online harms.

CCNC-SJ makes the following demands:

1. Defining hate speech
 - The Canadian government seek adequate community input to create a clear, robust and thorough definition of hate speech.
2. Timelines for removing hate speech
 - The Canadian government adopt timeframe requirements for social media companies to remove harmful material that are similar to those in existing and proposed legislation in the EU.
3. Identifying and removing (and not removing) content
 - An adequate identification and notification system be put in place should the government adopt a notice and action mechanism for identifying and removing harmful content.
 - More funding be provided to equity groups developing online tools to identify and flag harmful content.
4. Fringe cases
 - A specialized body such as a tribunal be created to expeditiously make determinations regarding fringe cases of online hate, and equity organizations have an opportunity to define hate speech and be involved in the creation of such bodies.
5. Penalties
 - Penalties be sufficiently high so as to discourage social media companies from allowing harmful material on their platforms.
 - Some civil recourse be provided to communities and individuals for harms caused by online hate speech or fake news.

Introduction: The harms of social media

It is well accepted that social media companies are not passive actors in the spread of hate speech and fake news online. Social media companies acknowledge that their

algorithms rank, sort and curate what users see in order to drive up engagements with their platform.¹

With hate speech and fake news drawing significantly more engagements due to its provocative and inflammatory nature, social media companies are profiting off this sort of hateful material.²

The societal harms and the corporate benefits from the promotion of hateful content are very real even if social media companies may not be intentionally promoting hate speech and fake news. Furthermore, the concerns expressed by these companies have little credibility when they take slow or ineffective action to curb the harmful content.

In a time when the Asian-Canadians and other communities are increasingly experiencing the widespread harm of online hate and disinformation, the public is becoming more aware of the major and direct involvement of social media companies in the spread of this harmful content. To this end, CCNC-SJ supports the Canadian government's decision to regulate social media companies as they relate to harmful content online.

The current state of the law

Unfortunately, despite enhanced scrutiny, legislative reform in Canada continues to fail to acknowledge the modern realities of social media's role in disseminating information.

Ostensibly as it stands now, Canada may be prohibited by American legislation from categorizing social media companies as publishers. As a signatory of the United States – Mexico – Canada Agreement in 2018, Canada may be obliged to implement s. 230 of the United States legislation, the *US Communications Decency Act*, which explicitly grants internet companies immunity from being held as publishers.³

Moreover, *Bill C-10*, announced on November 3, 2020, if passed, will amend the *Broadcasting Act* to add liability provisions for online broadcasting, but explicitly adds carveouts to ensure that social media platforms are NOT liable for posts made by its users.⁴ Why such a proposal has suddenly been tabled is disappointing, given the

¹ Facebook Company, "News Feed Ranking in Three Minutes Flat", (22 May 2018), online: *About Facebook* <<https://about.fb.com/news/2018/05/inside-feed-news-feed-ranking/>>.

² Peter Dizikes, "Study: On Twitter, false news travels faster than true stories", online: *MIT News | Massachusetts Institute of Technology* <<https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>>.

³ "Online platforms must be made liable for third-party hate content — and it might happen soon", (30 October 2020), online: *thestar.com* <<https://www.thestar.com/business/opinion/2020/10/30/online-platforms-must-be-made-liable-for-third-party-hate-content.html>>.

⁴ Friends of Canadian Broadcasting, "On the Proposed Amendments to the Broadcasting Act", (4 November 2020), online: <<https://friends.ca/explore/article/on-the-proposed-amendments-to-the-broadcasting-act/>>.

increased attention to the harms caused by social media, with some organizations suggesting it is the result of pressure and lobbying from large social media companies.⁵

Under ss. 318 to 320 of the Canadian *Criminal Code*, individuals can be liable for wilfully inciting hatred and genocide against an “identifiable group”, which encapsulates people “distinguished by colour, race, religion, national or ethnic origin, age, sex, sexual orientation, gender identity or expression, or mental or physical disability”. This incitement may also trigger penalties under the *Criminal Code* when the incitement of hatred and genocide is made online, and the *Code* empowers a judge to remove online hate propaganda.⁶

However, the *Criminal Code* fails to adequately protect people from the spread of online hate perpetuated by social media companies, because the provisions require that the incitement of hatred be wilful. This is a high bar to prove and has resulted in very few cases being pursued, and as such, very few convictions.⁷

Moreover, the *Criminal Code* provision may not apply to social media companies because the reason social media companies’ algorithms spread online hate is because it drives up interactions, not because they want people to be incited to hate or kill people of an identifiable group.

Given the current state of the law, it is imperative that Canada implements legislative reform to hold social media companies accountable for spreading hate speech and fake news online.

Requirements for effective anti-hate speech and fake news legislation

Defining hate speech

It is imperative that what constitutes hate speech is properly defined for this proposed legislation to work effectively and avoid public rebuke.

As noted, the anti-hate speech provisions of the *Criminal Code* do not adequately protect individuals from online harm, and likely do not cover social media companies.

It is therefore proposed that a definition of hate speech that does not have an intent requirement (similar to the now repealed section 13 of the *CHRA*) be enacted,⁸ or

⁵ Sarah Andrews, “Update to the Broadcasting Act: The Federal government surrenders to pressure from Netflix et al.”, online: *Friends of Canadian Broadcasting* <<https://friends.ca/explore/article/update-to-the-broadcasting-act-the-federal-government-surrenders-to-pressure-from-netflix-et-al/>>.

⁶ *Criminal Code*, RSC, 1985, c C-46, ss 318–320.

⁷ Richard Moon, *Submission to the Canadian Human Rights Commission concerning section 13 of the Canadian Human Rights Act and the regulation of hate speech on the internet prepared by Richard Moon* (Ontario Human Rights Commission, 2008) at 15–16.

⁸ Richard Moon, *supra* note 7.

alternatively, some definition that recognizes that social media companies can spread online hate without intending to persuade others to adopt those hateful messages.

The definition of hate speech must also be sufficiently robust, so as to address the effect that fake news can have on the standing and reputation of racialized communities and individuals.⁹

At the same time, it is imperative that the definition of hate speech be clearly defined, so as to avoid being so overbroad as to unduly infringe on freedom of speech. This will ensure that social media companies are able to properly comply with the legislation, as well as to avoid causing public rebuke of the legislation proposed, which is what happened with Germany's anti-hate speech social media law; the Network Enforcement Act.¹⁰

CCNC-SJ therefore stresses the need for the Canadian government to seek adequate community input to create a clear, robust and thorough definition of hate speech.

The process for removing (and not removing) material

From our review of other countries' existing and proposed legislation governing social media companies, several logistical, but crucial processes must also be clearly determined in order for such legislation to be effective. These processes are:

- a) How quickly will harmful material have to be taken down?
- b) How will offensive content be identified efficiently to social media companies?
- c) How will fringe cases be resolved?
- d) How will the quantum of penalties be determined and enforced?

a) How quickly will harmful material have to be taken down?

It is well established that the defamatory and inflammatory effect of hate speech and fake news is exacerbated the longer the content is available, as it will reach more people. As such, it is crucial that harmful material is removed expeditiously.

While CCNC-SJ has no exact position on what amount of time is reasonable, save that the sooner the better, we note that Germany's Network Enforcement Act¹¹ and the European Union's Code of Conduct on Illegal Online Speech requires harmful content to be removed by social media companies within 24 hours of notification.¹²

⁹ "China coronavirus: Misinformation spreads online about origin and scale", *BBC News* (30 January 2020), online: <<https://www.bbc.com/news/blogs-trending-51271037>>.

¹⁰ "Germany: Flawed Social Media Law", (14 February 2018), online: *Human Rights Watch* <<https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law>>.

¹¹ "Germany starts enforcing hate speech law", *BBC News* (1 January 2018), online: <<https://www.bbc.com/news/technology-42510868>>.

¹² European Commission, *European Commission and IT Companies announce Code of Conduct on illegal online hate speech* (2016).

b) How will offensive content be identified efficiently to social media companies?

Current proposals and legislation on social media legislation rely heavily on “notice and action” mechanisms which initiate content removal processes once social media companies are notified that harmful or illegal content has been identified. Examples of this model include the European Union’s Code of Conduct on Illegal Online Speech and the European Union *Digital Services Act*.¹³

The reason that this notice and action system has been adopted is likely because social media companies claim that they cannot adequately monitor and prevent the dissemination of all hate speech and fake news due to the sheer volume of social media posts made by users each day.¹⁴

However, notice and action systems are deeply flawed in its current state, because it puts the burden of identifying harmful material on users and or the government, rather than the social media companies who are best positioned to identify this content. Not only would it be extremely expensive for private individuals or the government to monitor all social media posts constantly for harmful content, it would likely be ineffective, particularly as social media companies’ algorithms may not be presenting the harmful content to those parties who are attempting to identify and report that harmful content. Coupled with the time-sensitive nature of removing harmful content, the notice and action system is not well-suited presently for regulating social media companies by itself.

A notice and action system can only be effective if it merely supplements adequate tools for identifying harmful content online. These tools may take the form of social media companies innovating or developing their algorithms to better understand what their algorithms are actually presenting to people, or alternatively that third-party tools and algorithms be developed to trawl through social media platforms to identify and report harmful material on a scale that people cannot.¹⁵

CCNC-SJ is currently involved in developing such a third-party tool. The tool in its present form works to detect online hate speech against Chinese Canadians in light of COVID-19 and will give users the option to hide or learn more about why a particular online post constitutes hate speech or misinformation. If the government is intent on using the notice and action system despite its serious limitations, then more funding must be provided to organizations like CCNC-SJ and the YWCA to develop such third-party tools to properly identify and report online hate.

¹³ European Commission, *Digital Services Act package: deepening the Internal Market and clarifying responsibilities for digital services* (2020).

¹⁴ Friends of Canadian Broadcasting, “Platform for Harm: Internet intermediary liability in Canadian law” (2020), online: <<https://friends.ca/workspace/uploads/documents/platform-for-harm-2020-friends.pdf>> at 14.

¹⁵ Juan Carlos Pereira-Kohatsu et al, “Detecting and Monitoring Hate Speech in Twitter”, (26 October 2019), online: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6864473/>>.

c) *How will fringe cases be resolved?*

It is inevitable that cases will arise where it is unclear whether the speech rises to the level of hate speech. For example, Bryan Adams was widely criticized for blaming “bat eating, wet market animal selling, virus making greedy bastards” for the Coronavirus outbreak, which for many was a clear snub at people of Chinese or Asian descent, and perpetuating the negative sentiment that Asian people were responsible for the Coronavirus outbreak. While Adams apologized, he claimed that his tweet was not aimed at a particular ethnic group, but was a promotion of veganism.¹⁶ This excuse has not been accepted by many, as it ignores the allusions to Asian people, such as the fact that wet markets are a common cultural practice in Asia, as well as with conspiracy theories that the Coronavirus is a result of Asian people’s dietary habits¹⁷ or was engineered by Chinese people as a biological weapon.¹⁸

As the Bryan Adams case exemplifies, it may be likely, perhaps even common, for fringe cases of hate speech to occur, particularly when the hate speech is made by persons who fear the repercussions of their actions, or do not realize they are harboring racist views.

CCNC-SJ takes the position that where fringe cases arise, social media companies should err on the side of removing the impugned material due to the time-sensitive nature of avoiding harms caused by hate speech and fake news.

To compensate for this, it will therefore be imperative that some process is created for determining or contesting whether impugned speech arises to the level of hate speech. This process must be accessible to either one of, or both, social media companies (who wish to receive a determination of whether content is hate speech or other illegal material that must be removed promptly) and private individuals (who claim their freedom of speech has been unfairly infringed and wish to seek a reversal of the decision to remove their content).

It is recommended that a specialized adjudicative body such as a tribunal be created that is empowered to decide on such fringe cases in an expeditious manner.

To be effective, it will be imperative that the adjudicative body has sufficient guidance as to what content may rise to the level of hate speech. This guidance must be robust and flexible and should be created with input from various stakeholders. While these recommendations are not exhaustive, CCNC-SJ recommends that factors like the hallmarks of hate determined in the case of *Warman v Kouba*¹⁹ be adopted:

¹⁶ “Singer Bryan Adams slammed as racist for post blaming ‘bat eating’ people for coronavirus”, online: <<https://www.nbcnews.com/pop-culture/pop-culture-news/singer-bryan-adams-slammed-racist-post-blaming-bat-eating-people-n1205166>>.

¹⁷ note 9.

¹⁸ “Coronavirus: US and China trade conspiracy theories”, *BBC News* (25 April 2020), online: <<https://www.bbc.com/news/world-52224331>>.

¹⁹ *Warman v Kouba*, [2006] CHR D No 50 .

- a) The targeted group is portrayed as a powerful menace that is taking control of the major institutions in society and depriving others of their livelihoods, safety, freedom of speech and general well-being;
- b) The messages use ‘true stories’, news reports, pictures and references from purportedly reputable sources to make negative generalization about the targeted group;
- c) The targeted group is portrayed as preying upon children, the aged, the vulnerable, etc.;
- d) The targeted group is blamed for the current problems in society and the world;
- e) The targeted group is portrayed as dangerous or violent by nature;
- f) The messages convey the idea that members of the targeted group are devoid of any redeeming qualities and are innately evil;
- g) The messages communicate the idea that nothing but the banishment, segregation or eradication of this group of people will save others from the harm being done by this group;
- h) The targeted group is de-humanized through comparisons to and associations with animals, vermin, excrement, and other noxious substances;
- i) Highly inflammatory and derogatory language is used in the messages to create a tone of extreme hatred and contempt;
- j) The messages trivialize or celebrate past persecution or tragedy involving members of the targeted group; and
- k) the messages contain “Calls to take violent action against the targeted group”.

CCNC-SJ again stresses the importance of community input for determining what constitutes hate speech.

Moreover, should a specialized body such as a tribunal be created to make determinations of what constitutes hate speech, CCNC-SJ stresses the importance that community and equity organizations like CCNC-SJ play a role in presiding over those tribunals.

d) How will the quantum of penalties be determined and enforced?

Adequate penalties must be determined to properly deter social media companies from allowing harmful content to be disseminated on their platforms. Without sufficiently large penalties, social media companies, particularly companies like Parler whose business model depends on allowing harmful material on their platform²⁰, may deem these regulatory fines simply as the cost of doing business, rather than seek to address and remove harmful content.²¹

Given the complex corporate structure of these large international social media companies, many of which are based in the United States, it will also be necessary for

²⁰ “Parler, a Twitter Alternative for Conservatives, Shatters Download Record Post-Election”, (9 November 2020), online: *MSN* <<https://www.thewrap.com/parler-twitter-alternative-conservatives-shatters-download-record/>>.

²¹ Friends of Canadian Broadcasting, *supra* note 14 at 45.

the Canadian government to ensure that any order for damages or regulatory penalty can be adequately enforced against social media companies.

Finally, there must be some guidance as to whether private individuals or communities may be able to seek redress for harms caused by the propagation of harmful material by social media companies.

Currently, there is no proven method for seeking private redress from social media companies for harms caused by online hate and fake news. It would therefore be prudent to clarify in this legislation whether, like the Manitoba *Intimate Image Protection Act*, assistance will be provided to affected individuals to help them seek redress. This may be in the form of civil recourse (such as pursuing a defamation action against social media companies by proving social media companies are publishers²², or seeking recourse under the *Canadian Human Rights Act*)²³, or creating a new statutory cause of action to allow civil court actions.

Finally, CCNC-SJ recommends any regulatory fines must be sufficiently high so as to encourage social media companies to innovate to prevent their algorithms to disseminate hateful material. Furthermore, at least some portion of these fines should be used to redress the harm caused to affected individuals and communities. This can be achieved either by directing some portion of the regulatory fines as compensation to affected parties or have that sum directed towards appropriate NGO's or government funded awareness and reconciliation projects.

Conclusion: Summation of Recommendations

With official recognition and empirical evidence of the social problems caused by online hate and inflammatory fake news targeted at Asian Canadians and other communities, CCNC-SJ has the following demands:

1. Defining hate speech
 - The Canadian government seek adequate community input to create a clear, robust and thorough definition of hate speech.
2. Timelines for removing hate speech
 - The Canadian government adopt timeframe requirements for social media companies to remove harmful material that are similar to those in existing and proposed legislation in the EU.
3. Identifying and removing (and not removing) content
 - An adequate identification and notification system be put in place should the government adopt a notice and action mechanism for identifying and removing harmful content.
 - More funding be provided to equity groups developing online tools to identify and flag harmful content.

²² Friends of Canadian Broadcasting, *supra* note 14.

²³ Richard Moon, *supra* note 7 at 5.

4. Fringe cases

- A specialized body such as a tribunal be created to expeditiously make determinations regarding fringe cases of online hate, and equity organizations have an opportunity to define hate speech and be involved in the creation of such bodies.

5. Penalties

- Penalties be sufficiently high so as to discourage social media companies from allowing harmful material on their platforms.
- The Canadian government must ensure that regulatory fines or other penalties are enforceable against large social media companies that can have complex corporate structures.
- Some civil recourse be provided to communities and individuals for harms caused by online hate speech or fake news.

Contact:

ryanc@ccncsj.ca

Ryan Chan

For media contact:

national@ccncsj.ca

<https://ccncsj.ca/>

Victoria Yeung at 647-613-0435