TekSavvy's Submission to the Department of Canadian Heritage Regarding its Consultation on Internet Harms

I. Introduction

TekSavvy Solutions Inc. ("TekSavvy") is pleased to submit the following comments in response to the Government of Canada's consultation on its proposed approach to address harmful content online (the "Consultation"). TekSavvy is an independent internet service provider ("ISP") based in Chatham, Ontario, and Gatineau, Quebec. It is Canada's largest independent ISP with a network across Canada and has been providing Canadian consumers with wireline broadband internet services since 2002. In addition to residential and business internet, TekSavvy also offers other telecommunications services such as telephone services and Internet Protocol television through its affiliate, Hastings Cable Vision.

Recognizing that there are public interest groups and researchers with expertise in these types of content and their regulation, TekSavvy seeks to make comments specifically from the narrow perspective of an ISP without commenting on all the broader considerations of the proposal. However, TekSavvy wishes to note that as a long-time champion of net neutrality and freedom of expression, it shares some of the concerns that others have expressed with the proposal. These include that the Consultation does not provide precise definitions for the types of harm or the types of electronic service providers to which it would apply; that it does not reference existing bodies of research on regulating the categories of online harms it addresses and that it is not effective to address all of these varied types of harms in a single piece of legislation.

From the perspective of an ISP, TekSavvy makes these submissions on the following topics:

- Since content moderation as proposed in the consultation would require automated systems, the framework should build on lessons learned from issues with other applications of automated systems for regulatory compliance;

- Since content moderation using automated systems at scale will necessarily involve automated decision-making, the framework should rely on or incorporate the Government's existing tools to evaluate automated and artificial intelligence systems employed for regulatory compliance; and,

- If site-blocking court orders are included in the legislative framework, there should be a clear set of factors or criteria that the court would be required to consider and weigh before issuing an order.

II. Use of Automated Systems for Regulatory Compliance

Automated processes will lead to over-censorship

The proposed 24-hour window for content moderation decisions, including decisions that would render the flagged content inaccessible in Canada, will almost certainly necessitate the use of automated decision-making. While this window may be appropriate for some types of illegal

content, we expect that applying this window uniformly across all types listed in the Consultation would result in a large degree of over-censorship.

A 24-hour moderation window would appear more achievable and less prone to error for some types of content subject to the proposal, such as child sexual exploitation material or intimate images that are found to be non-consensually shared. First, there are existing methods for automatically identifying copies of images known to fit within these categories. Further, this type of content would appear much less subject to nuance than the other listed categories.

For other types of prohibited content, such as hate speech and content inciting violence, however, there is much more room for the nuances of humour, sarcasm, fair comment, *etc*. —all of which can be expected to be difficult for an automated system to perceive or assess. Litigation that turns on whether something falls into these very categories can result in long court proceedings that reach the Supreme Court of Canada[1]; one could imagine the difficulty therefore in creating an automated system for assessing this type of content with any precision.

As a result, implementing a 24-hour window for making decisions on accessibility for all identified types of content would almost certainly lead to over-censorship. Platforms expected to meet this timeline would have every incentive to allow their automated tools to err on the side of making impugned content inaccessible in Canada in order to avoid incurring penalties for not meeting their regulatory obligations. Put another way, platforms would have every incentive to over-censor and little incentive to carefully consider the legality of content that may be on the fringes.

TekSavvy has experience with the use of automated tools for regulatory compliance through our development of systems for receiving and processing notices of infringement. Under the "notice-and-notice" provisions in sections 41.25 and 41.26 of the *Copyright Act*, [2] ISPs such as TekSavvy are required to forward a notice of infringement from a copyright owner to the subscriber at the IP address listed in the notice "as soon as feasible" once received. Copyright owners are prohibited from including some content, such as demands for payment or personal information, in their notices. The *Copyright Act* currently provides for statutory damages of not less than $5,000 and not exceeding $10,000 in the event that an ISP fails to perform its obligations. There is no express provision in the *Copyright Act* allowing for a due diligence defence, with the result that it is unclear if every failure to forward a notice would result in a fine (even where the ISP can show its due diligence).[3]

---

1 See for example, *R v Keegstra*, [1990] 3 SCR 697 or *Saskatchewan (Human Rights Commission) v Whatcott*, 2013 SCC 11, [2013] 1 SCR 467, both of which discuss the appropriate meaning of "hatred" at some length.

2 *Copyright Act,* R.S.C., 1985, C-42.

3 We note that this issue may be litigated in the Federal Court as a result of claims of almost $400 million filed by several copyright holders against Bell Canada for alleged failures to forward copyright notices. See Federal Court Docket T-1062-21, *Millennium Funding, Inc et al v. Bell Canada et al.*

TekSavvy receives many thousands of these notices on a weekly or even daily basis. As a result, like many ISPs, TekSavvy has developed an automated system for processing these notices, identifying subscribers, forwarding notices to the identified subscriber, and retaining customer information as required. Manual review over all notices it receives would be impossible for TekSavvy. However, TekSavvy's automated process cannot ensure that it does not forward non-compliant notices. Notices containing prohibited content cannot be detected with precision; if TekSavvy used a process by which to flag notices with certain keywords, for example, this would lead to some compliant notices failing to be forwarded. Because of the monetary risk associated with failing to forward notices and the lack of an explicit due diligence defence, ISPs must err on the side of over-forwarding notices to ensure their own regulatory compliance. This means notices that contain prohibited settlement offers or demands for payment, which can be intimidating to customers, continue to be forwarded to customers. This situation is the direct result of the notice-and-notice framework requiring perfect compliance at a large scale concerning imperfectly defined standards.

This can be analogized to the case of platforms, who, in seeking to meet a required 24-hour moderation window, and without a robust due diligence defence, would almost certainly err on the side of over-censoring. While 24 hours may be appropriate for categories such as child sexual exploitation material and sexual images shared without consent, we encourage the use of a longer window of time for other forms of harmful content — such as content suspected of being hate speech, content inciting violence, or terrorist content — to allow platforms a more rigorous and thoughtful review. Further, we encourage the inclusion of an explicit due diligence defence with well-defined criteria. Platforms could show, for instance, that their automated system was developed with diligence and in good faith, continues to be monitored for needed updates, and that it does a reasonable job of meeting the regulatory requirements. For an instance of content that was missed by the system, the platform would then have the ability to explain the criteria of its system to provide a reasonable explanation for the error, if one existed. For example, it would be defensible for a platform to show that it did not use a given criterion in an automated decision-making process because of an internal finding that it led to high instances of over-censoring which outweighed the harms caused by a given category of content.

If not, the risks of over-censorship of legitimate content are real. It could for example result in over-censorship of content from vulnerable or marginalized groups — the very groups the Consultation is in part designed to protect. This content may attract more complaints or "flags" on those platforms simply because of its dissent from opinions of larger groups or because of more targeted attempts to silence certain content. Open and thoughtful discussions from these communities could also use many of the keywords that automated systems use as criteria for taking down content. Examples of these moderation errors include Facebook's deletion of a woman's social media post detailing an experience in which her sons' were called a racist epithet[4] or a Twitter user who took responsibility for reporting sex workers' social media

---

[4] Dwoskin, Elizabeth and Tracy Jan, *The Washington Post*, "A white man called her kids the n-word. Facebook stopped her from sharing it," 31 July 2017.

accounts until they were shut down.[5] Automated systems required to blindly rely on the number of times content is reported or the use of certain keywords, in order to meet a strict 24-hour window, therefore, can be expected to lead to over-censorship.

Using Automated Decision-Making for Nuanced Decisions

The large platforms that we understand the Consultation seeks to address have existing content moderation practices in place that generally already use automated decision-making processes. In seeking to move part of these existing content moderation practices into the regulated sphere, the Government ought to ensure that these automated decisions do not serve to exacerbate some of the very issues that the Consultation seeks to address. For example, algorithms have the potential to make decisions based on criteria that have potential unintended biases in a manner that is not transparent to the public. For example, several studies have indicated that artificial intelligence models for processing hate speech were more likely to flag tweets as offensive or hateful when they were written by African Americans.[6]

As a starting point, TekSavvy submits that engaging with the Government of Canada's own *Directive on Automated Decision-Making*[7] and Artificial Intelligence Impact Assessment tool[8] could be a requirement for platforms in developing automated tools for making content moderation decisions. Platforms could also be required to provide transparency as to the criteria used in their automated decision-making, whether to the public or to the Government.

III. Considerations for Site-Blocking Orders

The Consultation proposes to provide the proposed Digital Safety Commissioner of Canada with the power to apply to the Federal Court for an order to require Telecommunication Service Providers to block or filter access to a service that has repeatedly refused to remove child sexual exploitation and/or terrorist content. We are pleased with the qualification in the Discussion Guide of this potential tool as an "exceptional recourse," the proposed judicial oversight over such orders, and the limited application to providers with violations regarding two of the five types of content (child sexual exploitation and/or terrorist content) as opposed to all forms of content the Consultation intends to address.

TekSavvy is of the view that site-blocking as an enforcement tool is generally simultaneously overly broad (as a result of the real risk of blocking legitimate content) while also ineffective. The only form of site-blocking that has been used in Canada to date requires ISPs to block access to specific domain names by removing those domain names from the ISP's domain name system (DNS). This is trivial to circumvent by the use of an alternative DNS service, many of which are

---

5 Clark-Flory, Tracy, *Jezebel*, "A Troll's Attempt to Purge Porn Performers from Instagram," 17 April 2019.

6 Ghaffary, Shirin, *Vox,* "The algorithms that detect hate speech online are biased against black people," 15 August 2019.

7 Government of Canada, *Directive on Automated Decision-Making*, 1 April 2021.

8 Government of Canada, *Algorithmic Impact Assessment Tool*, 1 April 2021.

freely available and that many internet subscribers already use without the goal of circumventing site-blocking. Even more sophisticated forms of site-blocking can be easily circumvented by those with only a moderate level of technical knowledge. Presumably the very persons seeking to access this type of content would be those most motivated to research and employ the fairly simple means of circumventing site-blocking mechanisms.

With that said, should the Government determine that site-blocking repeatedly non-compliant platforms could in some circumstances be an effective enforcement tool in incentivizing those platforms to comply (rather than for the block's purported efficacy in blocking access to content), we suggest that such an extraordinary remedy should only be available where it outweighs countervailing interests. To evaluate when that is the case, the statutory scheme ought to include certain criteria that courts would be required to consider in issuing such orders. We would suggest that these criteria include:

- **Instrument of last resort.** As recognized above, the site-blocking injunction should truly be an "exceptional recourse." In seeking an injunction, the Government should be required to demonstrate that it has sought other avenues of enforcement in order to reserve site-blocking for only those limited cases where other attempts have been unsuccessful. Put another way, the court ought to be convinced that alternative and less onerous measures were not effective. This would help ensure that site-blocking orders do not become a default enforcement mechanism but are instead reserved for extraordinary cases. Given ISPs' obligations not to discriminate against any traffic as a result of section 36 of the *Telecommunications Act,* [9] it is important to take measures to ensure that site-blocking is only used as an instrument of last resort.

- **Balance of freedom of expression considerations.** The Court should weigh the public interest in access to the platform in question against the enforcement considerations. This should include consideration of the degree to which Canadians' access to and engagement with legitimate content would be affected. There may, for example, be international platforms that are outside the traditional enforcement reach of the Canadian Government (absent international cooperation) and that do not consider Canada an important jurisdiction relative to their total user base. Other enforcement efforts against such platforms may therefore not have worked and the platforms may not have taken steps to meet Canadian regulations as a result of the small size of the jurisdiction. However, the platforms may still have other self-moderation policies in place that simply do not meet the criteria of the Canadian regime. The court should consider the effect of Canadians' loss of access to a platform of this type against the severity of the platforms' non-compliance. As an analogy, news media company CNN took the decision to withdraw its social media presence in Australia owing to Australia's decision to impose liability for defamatory comments on Facebook pages.[10] This result deprives Australians of access to legitimate content to which they otherwise would have access, which may

---

9 *Telecommunications Act*, S.C. 1993, c.38.

10 *B&T Magazine*, "US News Giant CNN Restricts Access To Facebook Pages In Australia Following High Court Ruling," 29 September 2021.

seem to be an outsized effect compared to the seriousness of defamatory comments on Facebook pages.

- **Technical clarity.** The statutory scheme should be clear with respect to the technical type of blocking that is required of ISPs, rather than suggesting ISPs use whatever means necessary to block a website. For example, as described above, to date the only form of site-blocking in Canada has been DNS de-indexing. However, it is easily circumvented through various technical workarounds, including the use of alternative freely available DNS servers or VPNs. Despite these possible circumventions, the statute should be clear that ISPs are not required to take additional, more invasive blocking steps that impose even greater burdens on ISPs. For example, IP blocking can affect unrelated internet resources beyond the site the blocking is intended to affect, and will impose higher operational costs to implement, maintain and trouble-shoot. Blocking based on content or specific protocols would require deep packet inspection, an invasive and advanced type of monitoring network traffic that would impose highly burdensome monitoring requirements on ISPs, would require specific equipment, and entail violations of the privacy of customers. The statute should be clear which type of blocking is required of ISPs, to promote certainty for ISPs and avoid incurring liability for not taking all possible blocking steps.

- **Clarity as to which party has the obligation to update the list of blocked sites.** As noted above, site-blocking carries a real risk of blocking legitimate content and stifling freedom of expression. It also is ineffective when websites are simply able to reappear under new domains or IP addresses. As a result, any list of domains subject to a blocking order will almost certainly need to be revised for currency and accuracy. The statutory regime should clarify that ISPs are not responsible for maintaining the list or liable for any inadvertent blocking of legitimate content. Instead, the Government must take reasonable steps to ensure that the list remains accurate (*i.e.*, including seeking revised court orders) and to identify issues of inadvertent blocking.

- **Consideration of the burden imposed on the ISP.** The scheme should expressly require the court to consider the burden that the injunction would impose on the ISP, including the aggregate effect of the injunction together with any other site- and application- blocking injunctions in effect for that ISP, as well as the technical feasibility and effectiveness of the proposed blocking in addressing the infringement.

IV. Conclusion

TekSavvy appreciates the opportunity to provide its comments on the Consultation. TekSavvy would be in favour of distinct regimes that are specific to the types of content at issue, based on the existing body of research on enforcement of these types of illegal content, and which engage in detail with the definitions of the harms at issue as well as technical details that will have significant bearing on the success of the regime. TekSavvy also reiterates the importance of issuing site-blocking orders only where prescribed statutory criteria are met.

Given these concerns, TekSavvy believes that the current proposal needs to be further developed and that more processes to consult on future developments and refinements are

required. We hope for more opportunities to participate in providing feedback as the discussion of online harms advances.