# Your response

| **Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.** |
|---|
| *Is this a confidential response? (select as appropriate)* <br><br> No |

To combat serious forms of harmful online content, the Canadian Government is committed to creating a legislative and regulatory framework to combat serious forms of harmful online content to protect Canadians and hold social media platforms and other online services accountable for the content they host. The government's goal is to keep all Canadians, especially children and equity-deserving groups and communities, safe from online harms and to provide them with an online ecosystem in which they are free to express themselves fully and safely, while also upholding Canadians' freedom of expression.

To help inform the development of its legislation, the Government held a public consultation on its proposed approach to address harmful content online from July-September 2021. Responses indicated a preference for a risk-based approach, anchored by a duty to act responsibly. In 2022, further engagement on a risk-based approach took place across Canada and included an expert advisory group, a Citizen's Assembly on Democratic Expression, and 19 Minister roundtables.

The feedback gathered during the roundtables provided valuable information and insights on what Canadians and key stakeholders are hoping to see included in future Canadian legislation. Canadian Heritage is drawing upon the information collected during the roundtables in developing policy and legislation. The Minister of Canadian Heritage continues to work to table legislation that aims to protect children, communities, equity-deserving groups, and Canadians online as soon as possible.

| **Question 2: Can you identify factors which might indicate that a service is likely to attract child users?** |
|---|
| *Is this a confidential response? (select as appropriate)* <br><br> No |

**Question 2: Can you identify factors which might indicate that a service is likely to attract child users?**

At a general level, the Government of Canada observes that social media platforms are services that are likely to attract child users as children are much more digitally connected now than ever before. These platforms can be used to threaten, intimidate, bully, and harass people – or used to promote racist, anti-Semitic, Islamophobic, misogynist, and homophobic views that target communities, put people's safety at risk, including children and youths. Social media platforms are also used by predators and offenders to exploit children.

Children are regularly exposed to cyberbullying and content that incites self-harm, including content that encourages and demonstrates self-mutilation (e.g., "cutting"), disordered eating, and suicide. These kinds of content can have a significant, cumulative negative impact on a child's physical and mental health, development, self-esteem and relationships.

More specifically, between July and November 2022, the Minister of Canadian Heritage conducted 19 roundtable discussions on key elements of a legislative and regulatory framework on online safety. Participants cited the vulnerability of youth to disinformation content, online luring, and inappropriate sharing of intimate images as three areas of concern. The impact of online harm on youth's mental health was a key point of concern with many participants sharing their first-hand accounts with youth who were in distress due to harmful content. Participants also noted the challenges parents face in protecting youth.

**Question 3: What information do services have about the age of users on different platforms (including children)?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 4: How can services ensure that children cannot access a service, or a part of it?**

*Is this a confidential response? (select as appropriate)*

No

The Government of Canada has limited evidence on this matter, related only to select input in roundtable consultations in 2022. Participants in our roundtables discussed the importance of platforms having authentication and verification systems in place. They also noted the importance for verification systems for platforms that include sensitive content such as pornography sites due to the risk of children accessing these platforms.

**Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?**

*Is this a confidential response? (select as appropriate)*

Yes

(response redacted)

**Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?**

*Is this a confidential response? (select as appropriate)*

Yes

(response redacted)

**Question 8: How do services currently assess the risk of harm to children in the UK from content that is harmful to them?**

*Is this a confidential response? (select as appropriate)*

Yes

(response redacted)

**Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?**

*Is this a confidential response? (select as appropriate)*

Yes

(response redacted)

**Question 10: What are the governance, accountability and decision-making structures for child user and platform safety?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[1]**

*Is this a confidential response? (select as appropriate)*

N/A

---

[1] See A1.2 to A1.3 of the call for evidence for more information on the indicative list of harms to children.

## Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

*Is this a confidential response? (select as appropriate)*

No

Participants in the Minister of Canadian Heritage's roundtables discussed the challenges youth experience in identifying and reporting online hate due to the inaccessibility of content flagging systems and the discrete forms of harmful content. Educating and providing resource supports to youth were identified as important solutions to pursue to prepare for instances when they encounter online harm.

There was agreement on the importance of increasing digital media literacy education to ensure people have the tools and knowledge necessary to identify online harm. School curriculum reform was suggested as a potential option to directly reach youth.

## Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

*Is this a confidential response? (select as appropriate)*

No

In March 2022, the Government of Canada established an expert advisory group on online safety, mandated to provide the Minister of Canadian Heritage with advice on how to design the legislative and regulatory framework to address harmful content online and how to best incorporate the feedback received during the national consultation held from July to September 2021. Here is a summary of the input and the discussion on the topics of recourse and complaints mechanism.

Most experts expressed that Canadians need a way to voice their grievances regarding platform behaviour. Some insisted that victims must be given the tools to trigger the content removal process. They illustrated that the appeal process on many online services is very difficult for users to navigate. They explained that in many instances legitimate actors and protest movements get suppressed with no avenue for redress. As such, experts emphasized, obligations must require services to provide user-friendly appeal processes that are quick, transparent, and fully functional.

Some experts suggested that each regulated service be required to have their own ombudsperson as part of what it means to be a responsible business. They explained that such requirements would be no different than mandating companies to have chief privacy officers, as the EU General Data Protection Regulation requires. Some experts stated that an internal ombudsperson would need to be supplemented by an external ombudsperson

office independent from both Government and the regulated services. They explained that the office would be staffed with individuals who have the necessary expertise to make assessments about context-specific content. Other experts questioned whether a social media council would be an appropriate way to provide victims with such recourse.

Additionally, some experts stressed the importance of compelling regulated entities to provide special support for children, including easy reporting options and terms that make sense to them (i.e., understandable language). They also argued that reports from children should be automatically prioritized for review, and that consideration should be given on how to facilitate reporting from parents and others on the child's behalf.

The Expert Group remained divided over the need and design of an independent, external recourse mechanism for content moderation decisions. Some experts emphasized that the issue is not whether to impose an external body to make takedown decisions, but rather how best to support users who have nowhere else to turn. They explained that under the framework, services would have their own internal complaint and transparency mechanisms. However, they stated, victims often feel that the platforms do not listen to their complaints, that they are still being harmed, and that the content they are flagging is not being addressed. They explained that victims need a venue to express their concerns. Experts emphasized that this support can be achieved without creating a content removal authority. They stressed that creating an independent body to make takedown decisions would be a massive undertaking akin to creating an entirely new quasi-judicial system with major constitutional issues related to both federalism and Charter concerns. Other experts argued that users need more meaningful recourse and sufficient access to justice. These experts argued that there must be a body to make takedown decisions in order to effectively help victims. On this latter point, some experts suggested that there be a connection between the ombudsperson and the courts, where the former could bring claims on behalf of victims to the latter. Some experts suggested that the new legislation promote the development of the courts to conduct their business online to be efficient and provide timely solutions to victims.

Furthermore, participants at the roundtable on key elements of a legislative and regulatory framework on online safety discussed the following element for best practices regarding reporting and complaints mechanisms:
- A desire for greater resources designed specifically for victims of online harm such as stronger reporting systems for harmful content and an ombudsperson who could advocate on behalf of victims and handle anonymous complaints
- An urgency for better reporting systems for online content. A few participants recounted instances where they reported online content and no action was taken. A few participants described the limited options available to users for reporting

content occurring over a live feed online or through private messages on a platform.

- A desire for platforms to offer more accessible reporting and content flagging systems
- A desire for platforms to provide tools that use plain language and include easy-to-understand instructions for reporting content.
- The concerns with existing moderation and content reporting systems where reported content is left untouched by social media platforms and instances where content is taken down due to being misidentified as being inappropriate content.
- The challenges youth experience in identifying and reporting online hate due to the inaccessibility of content flagging systems and the discrete forms of harmful content.
- Education and resource support to youth as important solutions to pursue to prepare for instances when they encounter online harm (digital literacy)

Additionally, many participants were frustrated at the content moderation systems social media platforms currently use. Several participants mentioned how content reporting systems are slow to respond to requests and rely on AI systems rather than human interactions, which can result in important contextual information being overlooked by AI systems. Participants also raised concerns that current reporting systems are not accessible for marginalized communities who may not have the knowledge base or ability to navigate the reporting systems social media platforms have.

**Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.**

*Is this a confidential response? (select as appropriate)*

No

- In the Minister of Canadian Heritage's roundtable with the tech industry, representatives from online platforms discussed the various measures their respective platforms have put in place to protect users including content moderating processes, age specific content restrictions, age-appropriate design features, and specialized platforms for children. A few participants noted the importance of educating users on how platforms operate and were supportive of establishing further educational campaigns and resources to inform users on how the online space operates.

**Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?**

*Is this a confidential response? (select as appropriate)*

N/AN/A

**Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/AN/A

**Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?**

*Is this a confidential response? (select as appropriate)*

No

N/AN/A

**Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?**

*Is this a confidential response? (select as appropriate)*

No

Many experts from the Government of Canada expert advisory group on online safety agreed that it would be necessary to impose robust transparency obligations on regulated services regarding their content moderation practice. Experts detailed the type of information that should be included in transparency obligations, including, but not limited to: quantitative and qualitative data concerning platform users; risk assessments; mitigation measures; volume and type of illegal and harmful content; content moderation procedures and outcomes; content moderator profiles

Some experts raised that a "safety by design" framework could be difficult to reconcile with the protection of children. They also explained that more robust obligations would be necessary to deal with child sexual abuse material. They suggested two elements to help promote the protection of children under a risk-based scheme: 1) compel services to adopt existing hash matching technology to stop child sexual abuse material from resurfacing on their platforms; and 2) compel services to use content moderation tools to compensate for gaps in AI detection.

**Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 22: How are human moderators used to identify and assess content that is harmful to children?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 23: What training and support is or should be provided to moderators?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?**

*Is this a confidential response? (select as appropriate)*

[Please select]

**Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?**

N/A

**Question 25: In what instances is content that is harmful to children, that is in contravention of terms and conditions, removed from a service or the part of a service that children can access?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 26: What other mitigations do services currently have to protect children from harmful content?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 27: Where children attempt to circumvent mitigations in place on a service, what further systems and processes can a service put in place to protect children?**

*Is this a confidential response? (select as appropriate)*

[Please select]

N/A

**Question 28: Other than those covered above in this document (the call for evidence), are you aware of other measures available for mitigating the risk, and impact of, harm from content that is harmful to children?**

*Is this a confidential response? (select as appropriate)*

No

We have one input to share in this regard, a project funded by the Government of Canada's Digital Citizen Contribution Program by CIVIX called "CTRL-F": Digital Media Literacy. This program uses an evidence-based approach for teaching students source evaluation and fact-checking techniques using lateral reading skills.. Through this project, CIVIX:
produced ready-to-use lesson plans, video tutorials and practice activities;
• trained teachers to deliver the learning module through capacity-building workshops;
• engaged thousands of teachers in all provinces and territories to use the learning module with their students; and,
• implemented a research plan to evaluate the new module's efficacy and the impact of the teacher training, which included an intensive skills study among students and a broader program evaluation.

After completing the module, CTRL-F, students were significantly more capable of evaluating online information. They were more likely to use lateral reading skills, accurately assess the trustworthiness of information, and identify the agenda of advocacy groups.

Based on lessons learned from funding recipients and consultations with key stakeholders, the most useful information to mitigate the risk, and impact of harmful content would be:
- how to recognize forms of online harm (luring, bullying, grooming, etc.)
- how to properly report online harm and what to expect from the reporting process
- the effects of online harm on its victims

**Question 28: Other than those covered above in this document (the call for evidence), are you aware of other measures available for mitigating the risk, and impact of, harm from content that is harmful to children?**

- best practices on fact checking
- lateral reading and bias recognition
- identifying propaganda (understanding 'othering' and being presented with emotional arguments)
- building communication skills, especially constructive dialogue skills which assist in navigating challenging and polarizing topics.