

## ONLINE SAFETY ACTION TABLE – PROGRESS REPORT

September 2024

### BACKGROUND

The Premier’s Online Safety Action Table (OSAT) was created to take action to improve online safety for British Columbians – especially for our youth.

The OSAT provided a forum for discussion between the Attorney General, Google Canada (including YouTube), Meta, TikTok, Snapchat, and X (formerly Twitter).

The OSAT proceeded on an ambitious timeline. Several meetings involving the Province and all OSAT participants were held between May 10 and August 28, 2024. Representatives from the Province and participating companies also held frequent individual meetings to discuss issues in detail.

This report provides an overview of progress on agreed actions, outcomes, and next steps.

### OSAT ACTIONS AND OUTCOMES

The OSAT process served as a valuable forum to discuss a range of safety concerns with some of the largest technology companies offering products and services in British Columbia. While no formal meetings were scheduled beyond August 28, 2024, participants acknowledged a willingness to explore future opportunities in this area.

#### ***Emergency Information***

OSAT participants agreed on arrangements to amplify verified public safety information, including evacuation orders and alerts during emergencies such as wildfires. This work commenced during the 2024 wildfire season, providing more accurate information to B.C. residents.

The Province and several participating technology companies are continuing to develop and refine the implementation of these new arrangements. No problems or issues were reported during the summer of 2024.

Some companies also offered advertising credits to the Province to facilitate the relay of emergency preparedness information and resources to people through the platforms.

#### ***Non-Consensual Intimate Images***

A key priority of government is stopping the non-consensual distribution of intimate images on social media platforms, including cases where this harm is connected to sextortion.

These problems are especially acute for young people, where this form of sexualized violence can have devastating consequences.

OSAT participants agreed to implement new measures aimed at stopping the distribution of non-consensual intimate images on their platforms:

- Creation of a dedicated channel to streamline the implementation of orders made under B.C.’s *Intimate Images Protection Act*, and
- The development of a mechanism for the Ministry of Public Safety & Solicitor General’s Intimate Image Protection Service to act as a “trusted flagger” authorized to notify participating companies through a direct escalation channel when a non-consensual intimate image is reported by a victim so the company can take appropriate action.

The new processes allow people to quickly report non-consensual intimate images and improve enforcement of orders made under B.C.’s intimate image legislation.

As of August 28, participating companies had already reported using the new “trusted flagger” mechanism to address non-consensual intimate images.

### ***Protecting Children Online***

OSAT participants had extensive discussions on the topic of online safety for children. The challenges in this area are complex, and the risks and threats to the health, safety and well-being of children continue to evolve. Balancing the need to protect children with a desire to preserve their rights, including continued access to, and appropriate use of, social media, requires collective action and a combination of measures or strategies.

The Government of Canada has proposed federal legislation to address a wide scope of harms children face online. Several other jurisdictions are advancing “safety by design” initiatives to build protections directly into online products and services. Other jurisdictions have created oversight roles and access to individual redress for people who suffer harm online.

Litigation is another means to correct the harmful behaviours of commercial actors whose products and services are alleged to have harmed children. Lawsuits have been started against technology companies in multiple jurisdictions.

Education and awareness campaigns can be used to draw attention to the risks children face online and to equip them with the knowledge to protect themselves. These efforts range from platform and industry-led initiatives to broader statements or warnings by government bodies.

During the OSAT process, the Province emphasized the need for commercial actors to prioritize safety and to protect children who use their products and services. Although the technology companies provided comprehensive descriptions of their existing measures intended to keep children safe online, the effectiveness of these efforts remains uncertain in the absence of independent expert reviews supported by sufficient access to relevant data and evidence, much of which is held only by the companies themselves.

Despite industry-led efforts to promote safe online experiences and increase education and awareness, more work is necessary. Children continue to face a range of risks and harms in the online environment. British Columbians have witnessed the tragic loss of life of young people who were victimized by actors using online platforms as a medium to inflict serious harm. Children using these platforms experience a range of risks that often include mental health, anxiety, sextortion, and bullying.

During the OSAT process, the Province invited participating technology companies to voluntarily accept and sign a Declaration on Online Safety for Children (the Declaration), detailing a range of duties and specific protections aimed at keeping children safe online (summary attached as Appendix A). In addition to advancing a combination of guiding principles and specific safety features for platforms, the proposed Declaration included a commitment to provide information and evidence to an independent, expert evaluator to determine the effectiveness of various safety measures and to identify new risks and mitigation strategies for future work.

None of the technology companies participating in the OSAT process elected to sign the Province's proposed Declaration. However, some companies expressed a preference for the Province to legislate in this area so that child-safety requirements would apply equally to all technology companies offering similar types of products and services.

Some technology companies committed to implementing new unilateral measures to promote online safety for children, including actions that respond to, or that were informed by, the OSAT process. While companies advised that some of these changes have already been implemented, others will be announced in the coming weeks, including a range of product changes by Meta to promote online safety for children that are expected to be announced September 17, 2024. In the context of a voluntary initiative like OSAT, the efforts announced by the technology companies may contribute to the objective of improving online safety for children.

Attached as Appendix B is a list of actions that the technology companies report were influenced by the OSAT discussions.

Further expert analysis will help to understand the effectiveness of youth safety measures and to identify additional options for the future.

## **POTENTIAL NEXT STEPS**

The Province can continue efforts aimed at ensuring online platforms and services operated by technology companies remain available to deliver reliable government information in emergency situations like wildfires.

The Province can also continue to engage with participating technology companies who express a willingness to implement streamlined processes aimed at stopping non-consensual intimate images and support the objectives of the *Intimate Images Protection Act*.

In the area of online safety for children, multiple strategies and collective action will likely be required for the foreseeable future. The Province may therefore decide to explore any of the following options:

- Consultations with the public, experts and stakeholders to develop better understanding of the needs, interests, challenges and opportunities in this area.
- Legislation to:
  - Regulate the products and behaviour of technology companies who make their products and services available in B.C.,
  - Provide access to redress for individuals who suffer harm through the use of these products and services, and
  - Identify processes or standards to measure compliance and effectiveness of legislated online safety measures, including enforcement mechanisms.
- Creation of an authority to provide oversight and advocacy for online safety for children.
- Leading or supporting trustworthy, evidence-based public education and awareness campaigns raising awareness of risks and promoting online safety.
- Litigation to correct the harmful behaviour of actors in this area and to recover government's costs incurred responding to those harmful behaviours.
- Explore potential coalitions and collaborations with other governments to advance a broader collective response to the risks facing children online.
- Encourage further research and innovation among experts, post-secondary institutions, and industry.

## **CONCLUSION**

Working to an ambitious timeline and engaging directly with some of the largest technology companies who provide online products and services used by people in British Columbia helped to secure actions to improve online safety in this province – especially for youth. The improvements with respect to government information in emergencies and non-consensual intimate images are relatively clear. On the issue of protecting children online, technology companies reported that direct engagement influenced them to implement new safety measures and practices. The discussions also helped to illuminate the limitations and reliability of voluntary change, which will inform future decisions and options the Province may decide to pursue in the future.

## APPENDIX A

### Summary of Declaration on Online Safety for Children

#### *Duties to Protect Children*

The Declaration invited the technology companies to acknowledge:

- A duty to act responsibly and prioritize the best interests of children using their products or services.
- A duty to protect and safeguard children by eliminating or mitigating risks to their health, safety and well-being resulting from the use of their products and services.
- A commitment to not take actions that circumvent or frustrate the duties and standards listed in the Declaration.

#### *Age-Appropriate Applications*

The Declaration invited the technology companies to voluntarily commit to:

- Not make any products or services available to children online until it has undergone a documented process focused on identifying and eliminating or mitigating risks.
- Take appropriate measures to establish the age of users, balancing the risks to children using their products or services with the rights, freedoms and privacy interests of children.
- Advance clear, age-appropriate approaches to obtaining permissions or consent from children or parents or guardians acting on their behalf.

#### *Specific Safety Designs and Settings*

The Declaration invited the technology companies to voluntarily commit to:

- Apply default safety settings for children online in relation to account privacy, geolocation, personalized advertising, screen-time limits, alerts and notifications during sleep hours and persuasive design features that keep users engaged with the products or services.
- Provide age-appropriate online options to turn off default safety settings only if the company can demonstrate it would not create an inappropriate risk or harm to children, and if obvious indicators are provided that the setting is being changed.
- Avoid using recommender systems for children unless the company can demonstrate that their systems effectively restrict and filter content that is unlawful or harmful for children, and that mechanisms are in place to allow users to report and filter content they don't want to receive.

- Put age-appropriate restrictions on online content that can be discovered through the platform, or that is provided through advertisements.
- Report on the effectiveness of their online content and search moderation measures designed to promote online safety for children, including performance measures such as the time to respond to complaints and the accuracy of decisions concerning those complaints.

### *Reporting and Support*

The Declaration invited the technology companies to voluntarily commit to:

- Provide age-appropriate mechanisms to flag or report content and messages that violate the company’s terms of service or could pose a risk to the health, safety or well-being of children.
- Provide age-appropriate mechanisms specifically for the purpose of reporting non-consensual intimate images or sextortion, and timely interventions to address such reports, including blocking communications from wrongdoers and reporting a child in need of protection from sextortion to law enforcement.

### *Data and Privacy*

The Declaration invited the technology companies to voluntarily commit to:

- Not collect, use, share or retain a child’s personal information or other user data in ways that are detrimental to the child’s safety or best interests.
- Provide options to minimize the collection, use, sharing or retention of a child’s personal information or other data, while minimizing negative impacts to the child’s experience using their products and services.
- Ensure that any warnings or requests for permission to collect, use, share or retain personal data use simple language.
- Avoid the use of deceptive design patterns intended to encourage or manipulate children into sharing more personal information and data than they otherwise would have.

### *Education, Awareness and Supports*

The Province invited technology companies to identify, and commit to, education and awareness activities that would promote online safety for children.

### *Evaluation*

The Declaration invited the technology companies to voluntarily commit to:

- Provide evidence and information demonstrating compliance with the above measures to an independent evaluator, appointed by the Province, with expertise and experience in online safety for children.
- Provide non-identifiable user and engagement data to researchers and civil society groups to foster transparency and better understanding of the experiences of children online, and to develop strategies to protect them.

## APPENDIX B

### Actions Taken by Technology Companies

The companies who participated in the Online Safety Action Table (OSAT) process were invited to provide a summary of actions that were informed by, or responsive to, the OSAT discussions.

The following list, compiled by the Province in September 2024, is a synthesized inventory of actions identified by Google Canada (including YouTube), Meta, Snapchat, TikTok and X. Submissions from the companies have been edited for brevity and consistency.

#### EMERGENCY INFORMATION:

##### Google:

- Google has updated the local government resources that they have on file to ensure that they continue to serve reliable, accurate information when emergency alerts are issued on their products.
- They are engaged in ongoing conversations with the B.C. Emergency Management team to support the government to create and provide alerting information to directly integrate to Google’s Public Alerts system.

##### Meta:

- Meta has worked over the summer with B.C. officials to customize their emergency response tools to specific crises in B.C. (notably wildfires) and to ensure that those emergency response tools contain credible, accurate information for B.C. residents. In particular, Meta worked with officials to pin posts from government sources (Emergency Management B.C., DriveBC, B.C. Wildfire Service, impacted municipal governments) at the top of Meta’s Crisis Response Pages so that they are easily accessible and so that users have the latest information available to them. In addition, Meta distributed advertising credits to the B.C. government to assist with emergency response and deployment of credible information on an ongoing basis.

##### Snapchat:

- Snap has offered both the B.C. Government and their advertising agency ad credits/coupons to support the B.C. Government's emergency response efforts.

#### NON-CONSENSUAL INTIMATE IMAGES (NCII)

##### Google:

- While Google and YouTube already had existing reporting mechanisms for individuals and governments to report illegal content on their platforms, including content violative

of the B.C. *Intimate Images Protections Act*, Google provided the Attorney General's office with detailed outlines of the pathways that both victims and the government can use, and the responses they can expect to receive, when reporting content both with and without a Civil Resolution Tribunal Order.

- Google also made product change improvements to help people affected by non-consensual sexually [explicit fake content](#).

#### Meta:

- Meta worked with officials to establish new, dedicated reporting and escalation channels for NCII, both prior to and following intimate image protection orders from the Civil Resolution Tribunal. Since those processes were established this summer, Meta has received more than half-a-dozen pre-order "trusted flagger" reports from the Intimate Images Protection Service (IIPS) and have been able to quickly remove the content and accounts identified. These offending images will be included in Meta's existing processes and cross-platform programs designed to stop them from resurfacing online.

#### Snapchat:

- Snap worked with the B.C. Government to finalize a process for onboarding the IIPS as a Trusted Flagger.

#### TikTok:

- TikTok advocated to establish informal, "pre-order" reporting channels with platforms to remove NCII, before victims even seek out a Civil Resolution Tribunal order, and worked with officials to establish central channels to ensure that any orders will be received and actioned promptly. B.C. agencies are now being onboarded to TikTok's Community Partner Channel, which will allow for rapid removal of NCII when it is first identified, whether before or after an order is issued. Based on information provided by the Province, to date there have been no orders sought under the *Intimate Images Protection Act* involving the non-consensual distribution of intimate images. Images containing nudity or sexual abuse such as those used in image-based sexual abuse and sextortion on TikTok are broadly disallowed by TikTok's Community Guidelines.

#### X:

- In consultation with the IIPS of the Ministry of Public Safety and Solicitor General, X aligned on dedicated reporting pathways for intimate images on the platform. X doesn't yet have any indication or communication that the IIPS has had to utilize the dedicated reporting pathways.

## ONLINE SAFETY FOR CHILDREN

### Meta:

- Meta introduced Teen Accounts on Instagram, which is a new experience on Instagram for teens, guided by parents. Teen Accounts include built-in protections for teens, which will give parents peace of mind that their teens are using Instagram to speak to friends and explore their interests - without having to worry about unsafe or inappropriate experiences. These built-in protections limit who can contact them and the content they see - and have been designed to address parents' biggest concerns, including:
  - who their teens are talking to online,
  - the content they're seeing, and
  - whether their time is being well spent.

The Government of British Columbia provided feedback on this approach as part of Meta's participation in the province's Online Safety Action Table. The launch of Teen Accounts in British Columbia, and Canada more broadly, builds on this important collaboration between government and industry to help young people stay safe online.

Meta undertook several other activities related to youth safety:

- [Combating Financial Sextortion Scams](#)
- [Meta and Center for Open Science Open Request for Proposals for Research on Social Media and Youth Well-being Using Instagram Data](#)
- Prevent Alliance - In early July, Meta officially signed onto the Alliance to Prevent Drug Harms ("Prevent Alliance"). The Prevent Alliance brings together the Department of State, United Nations Office on Drugs and Crime (UNODC), and the technology industry to prevent the abuse of digital platforms for illicit and harmful activities related to the non-medical use of synthetic drugs.
- [Meta Joins Thorn and Industry Partners in New Generative AI Principles](#)
- [New Tools to Help Protect Against Sextortion and Intimate Image Abuse](#)

### Google:

- Google announced [School Time mode](#), to provide parents with more control over their kids' use of technology during school hours.
- YouTube launched [teen supervision tools](#) that enable parents to have more oversight/involvement.

- YouTube will be holding an event with MediaSmarts in the Greater Vancouver Area in November 2024, to engage educators, advocates and experts in a discussion around youth online safety.
- MediaSmarts and YouTube are also launching an educator resource in late October 2024 for teachers to use in classrooms. This educator workshop will provide young people with resources, tools and programming related to digital literacy skills.
- Google and YouTube will be running a public advertising campaign promoting youth online safety, which will be running in B.C. for the second time this year, launching in Q4.
- Google also plans to promote pre-existing educational resources that support the OSAT objectives in B.C.

#### Snapchat:

Since the OSAT was convened, Snap [announced](#) plans for new features to further safeguard the Snapchat community from online harms. While not specifically informed by the OSAT process, these expanded protections will address many of the concerns expressed by the B.C. Government during the discussions at the action table:

- Expanded in-app warnings: Teens will see a warning message if they receive a chat from someone in a region where the teen’s network isn’t typically located.
- Enhanced friending protections: When teens send or receive a friend request from someone they don’t have mutual friends with, and that person has a history of accessing Snapchat in locations often associated with scamming activity, Snapchat prevents delivery of the friend request altogether when Snapchat detects the person is a bad actor, resulting in the removal of that account.
- Simplified location-sharing and additional reminders: All users – including teens – now receive more frequent reminders to make sure that they are always up to date on which friends they’re sharing their location with on the Snap Map. Snapchat also introduced simplified location-sharing, making it easier for Snapchatters to customize which of their friends – if any – can see their location.
- Blocking improvements: Blocking a user now blocks new friend requests sent from other accounts created on the same device.

#### TikTok:

- As part of TikTok's commitment to youth safety, the company partnered with MediaSmarts to support the creation of *Talking TikTok: A Family Guide*, which provides important information to Canadian parents, guardians, caregivers and teachers to help ensure a secure and positive experience for their teens on TikTok:

[https://mediasmarts.ca/sites/default/files/2024-05/guide\\_talking\\_tiktok\\_family\\_guide.pdf](https://mediasmarts.ca/sites/default/files/2024-05/guide_talking_tiktok_family_guide.pdf)

X:

- In May, X officially joined [Project Lantern](#) to share and ingest signals with other participating companies to help root out cross-platform bad actors that are exploiting multiple services to traffic Child Sexual Abuse Material.
- In July, X [announced](#) their implementation of [Thorn's](#) text classifier to detect potentially harmful conversations — such as sextortion and solicitation of nudes — occurring between bad actors and minors.