Four things to watch out for in the Online Harms Act 2.0

Author: Taylor Owen, Founding Director of the Centre for Media, Technology and Democracy and Beaverbrook Chair in Media, Ethics and Communications

I have worked on online harms policy in Canada for the better part of a decade, on the federal expert panel advising this government and the last, and on the national AI Task Force. A redrafted Online Harms Act is expected this week. This is the third time the government has brought forward an approach to digital safety, and this time we are in a very different place. Public support for governance has never been higher, and a new wave of AI products has reinforced the need for guardrails.

What follows are four things to watch out for in the new bill. They draw on my submissions to the AI Task Force and the Expert Panel on Online Safety, and on the federal and provincial policy work that followed the Tumbler Ridge shooting. For each, I have tried to be specific about the design choice, the likely objection, and the risk to passage if it is handled badly.

1. Don't rewrite the architecture, amend it

The core framework in Part 1 of Bill C-63 was sound: a functional definition of regulated services, a duty to act responsibly, a duty to protect children, systemic risk assessment, transparency obligations, and an independent regulator. This is the same architecture every comparable democracy has adopted, in the UK, the EU, Australia, and Ireland. Canada does not need to invent a new model. It needs to pass the core of the one it already drafted.

Being last carries a compensating advantage. The UK, the EU, and Australia have been running these regimes for a few years now, and Canada can take their design choices, their early case law, and their enforcement record and write a sharper bill than any of them managed the first time out. It has reason to stay close to them as well. The same handful of companies are regulated in all of these markets at once, and a Canadian law whose obligations line up with its peers' is easier to enforce, cheaper to comply with, and harder to pick apart than one that invents its own categories. The point is not to copy. It is to improve on what already exists without drifting out of step with it, because divergence for its own sake helps no one but the firms looking for the gaps between regimes.

The bulk of the objections to C-63 were always about the other parts of the bill, the Criminal Code and hate-speech amendments that drew most of the controversy. Those were split off in December 2024 and are a separate debate. What remains is the platform safety regime, which is uncontroversial in design even where it is contested in detail.

The real cost now is delay. Canada is already the last of its peers to act. The bill should in my view reintroduce Part 1, with targeted amendments on: the under-16 pause and the age assurance it depends on, consumer-facing chatbots, the design of the regulator, the perimeter of services covered and the duties that attach to them, and a few specific harms still to be settled.

2. Pause under-16 access, don't ban

This is the most important design choice in front of the government, and the one that is the most fraught.

The government is exploring a ban on social media access for those under sixteen. Bans of this kind are blunt, easy to circumvent, and the most legally exposed instrument available. Australia's came into force in December, removed or restricted 4.7 million accounts in its first weeks, and is now facing a constitutional challenge from Reddit and from two teenaged plaintiffs.

But the impulse behind a ban is completely understandable. Parents are frustrated, because both the companies that build these products and the governments we task with protecting us have failed. Parents are watching products they did not design and cannot control shape their children's lives, and the answer they keep being given is to wait for a regulator that does not yet exist and may be two or three years from functioning. It is fair for them to ask that something be done in the meantime.

In my view the better way to respond to that demand is to treat the restriction as a moratorium rather than a ban. A ban is an end state. A moratorium is a mechanism: a temporary restriction on under-sixteen access that lifts once a platform proves, through its regulatory obligations, that its product is safe for young users. The difference matters. A ban asks nothing of the platform and leaves its design untouched. A moratorium makes reform the price of access, and it puts the burden where it belongs, on the companies to demonstrate safety rather than on young people to stay out. It is also more proportionate, and so more defensible, because it targets the product features that cause harm rather than a young person's right to be online.

What would a platform have to prove? The obligations are for the regulator to set, but the central one is compliance with an Age-Appropriate Design Code. In practice that should reach the design itself: no algorithmic amplification into minor accounts, compulsive-use features like autoplay, infinite scroll, and streaks turned off for minors, and independent audits of recommendation and age assurance systems. The restriction lifts when those conditions are met.

Age assurance is the most difficult part of any online safety law, since the system you build to check ages is itself a privacy challenge. There are no clean answers to age verification, but there are choices that can be made at the legislative stage to enable the best possible implementation, and to learn from it once it is in operation.

Start with the technology's limits. Australia's age assurance trial found the best facial age-estimation systems off by 1.3 to 1.5 years on average, with accuracy slipping near the age threshold and for women and non-white users, and Indigenous users underrepresented in the data the systems were trained on. A tool that mistakes a sixteen-year-old for fourteen or eighteen cannot carry the full weight of a hard governance line. And Canada's privacy law still offers no specific protection for children's data. As written, an age check would require young people to hand over sensitive information to comply with a rule that does nothing to protect that information.

So it is critical to amend PIPEDA for youth data ahead of or alongside the Act. It is also important to prohibit photo-ID uploads as a condition of age assurance, and separate the verification function from the service wherever possible. The EU's age-verification blueprint shows this is feasible: zero-knowledge proofs can confirm that someone is over a given age without revealing who they are.

Two cautions. Be skeptical of broad parental-consent exceptions. The American experience shows they are easily circumvented and fall hardest on children in unsafe homes. And build for the young people for whom online spaces are protective rather than only risky. Service-specific exceptions for crisis lines, Indigenous cultural platforms, and mental-health services should be in the design from the start.

3. Include chatbots, but with their own obligations

The biggest gap in the previous bill is that it didn't address consumer-facing AI chatbots. While not without complication, in my view they belong inside the framework. The combination of risk assessments, risk mitigation reports, transparency requirements and age appropriate design will also work for consumer facing AI. But they are a different kind of product, and they pose a different kind of risk, so the obligations attached to them should be bespoke to chatbots rather than borrowed from the rules written for social media.

The difference is in how the harm happens. A social media harm is mostly about distribution, where content that already exists reaches someone it should not. A chatbot harm is about generation, where the system produces the content directly, in a private exchange that simulates intimacy and trust. We have already seen what that can produce. OpenAI's own systems flagged the conversations of the person who went on to kill eight people in Tumbler Ridge, employees argued internally for referring the case, leadership decided it did not meet a threshold, and Canadians learned about it from a newspaper. xAI's Grok was generating sexualized images of women and, in some cases, children, with regulators in France and the UK moving on it and none in Ottawa. Character.AI and Google agreed to settle a suit brought after a fourteen-year-old's months-long relationship with a chatbot ended in suicide. In each case, no Canadian authority had any standing to ask how the system was built or why it failed.

The objection to regulating chatbots is that they are a general-purpose technology and fast-moving, and that regulating them will chill a nascent technology. The answer is to be precise about what is regulated. Obligations should be attached to consumer-facing deployment and to a defined set of serious harms, including self-harm and suicide, child sexual abuse material, non-consensual intimate imagery, and credible threats of violence. They should not attach to research or model development. In other words, regulate the harness, not the model.

In practice that means a functional definition that captures both standalone chatbots and embedded features like Meta AI or Grok inside X, mandated safety protocols for those harm categories rather than internal company thresholds, and child-specific design rules, which could include no simulated romantic or sexual interaction with minors, and limits on persistent memory and engagement-maximizing features for minor accounts.

Writing the detail of those obligations is the regulator's job, and they will need to differ from the social media versions. For example, transparency. A social media transparency report is mostly about distribution at scale: what content was removed, how fast, and how the recommender was tuned. A chatbot report is about the model's behaviour: what it was tested against before release, how often it failed those tests, and what it did when an exchange crossed into one of the serious-harm categories. Or Age-Appropriate Design Code. For a feed, the design questions are amplification, autoplay, and infinite scroll. For a chatbot they are the interaction itself, whether it simulates romance with a minor, how much it remembers between sessions, and whether it breaks contact and points to help when a young user signals distress. The statutory duty can be the same. What it requires in practice is not, because the product is not.

4. Build the regulator that shows government can innovate

The original bill got the most important institutional choice right. It called for a new, independent regulator, which is the only serious option. The obligations in this framework, the risk assessments, audits, age assurance testing, and transparency reporting, cannot be run out of a department or bolted onto an existing body with a different mandate and culture. They need a dedicated regulator with the independence and capacity to act on them.

The bill also set out how large that regulator should be, with a costing built around roughly 330 staff at maturity. What it said almost nothing about was the harder and more important question: how it would get the right people, and how it would work. Size matters, but mainly in service of the real test, which is whether the body can enforce the obligations and keep pace with the companies it oversees. That takes two things the public service does not reliably produce. The first is deep technical capacity, people who understand recommender systems, model behaviour, and age assurance well enough to test a platform's claims rather than take them on faith. Ofcom, preparing to enforce the UK law, hired around 350 people for online safety and recruited a number of them directly from senior roles at Meta, Microsoft, and Google. That is the calibre of

hire this work requires, and it is not the kind of hire the standard federal classification and pay structure is built to make. The second is nimbleness, the ability to move closer to the speed of the technology than the speed of a normal government process.

A caution on speed, though, because the case for nimbleness is easy to overstate. The goal is not to match the industry's pace for its own sake. Part of what good governance does is introduce friction, a check on decisions that today carry none, a requirement to weigh a risk before a product ships rather than after a harm is done. The companies will call that friction a cost, and some of it is, but it can also be the point. The craft is to put friction where it slows harm and to strip it out where it only slows the regulator.

This is also an opportunity. Canadians are increasingly skeptical that the government can build things that work. A regulator that is fast, technically credible, and visibly effective would be a direct answer to that skepticism. But it will not happen on the standard template, because an institution built the usual way inherits the usual constraints. To be different, it has to be built from the ground up to be different.

Concretely, that means a standalone body rather than a unit inside an existing regulator or department, which carries neither the culture nor the capacity for this work. It means authority to hire outside the standard Treasury Board classifications, perhaps on the model used for the Bank of Canada, so that recruiting the right technical people is possible rather than aspirational. It means offices across the country rather than in Ottawa, close to the talent, the companies, and the communities the regulator serves. And it means setting the size as a floor to build from, not a ceiling, with a statutory review that scales it as the mandate grows.

Capacity is only half of it. A regulator can only govern what it can see, and today the companies hold nearly all the information about how their systems work. The bill must close that gap, and it should be deliberate about who gets what. Independent vetted researchers should be able to reach the underlying platform data needed to study systemic risks, on the model of the EU's Article 40 regime, whose data-access portal opened in October. The regulator itself needs something narrower and more routine: a limited form of mandatory reporting on the companies' own safety decisions, when their systems flagged something and against what thresholds, and when they decided not to act, without the regulator taking in the content of private interactions. Researchers get depth, the regulator gets the pattern.

How both of these should work for chatbots is still an open question, and a different one than for social media. The unit is not a public feed but a private one-to-one exchange, which changes what underlying data means for a researcher and what a useful flagging report looks like for the regulator. The bill does not have to settle it now. It has to create the powers and direct the regulator to work them out.

Conclusion

While we are behind as a country on this issue, there is a more hopeful way to read the years of delay: Canada is late, but late means the hard lessons have been learned elsewhere first. The UK, the EU, and Australia have built these regimes and shown what works and what to avoid. The model is no longer speculative.

That is the opportunity, and it runs deeper than the bill itself. Few governments get to build a regulator from scratch, and fewer still at a moment when so many Canadians have given up on the idea that the state can build anything new that works. Done well, this is more than a safety regime. It is the answer to that doubt, proof that the country can still build something new and make it work, a regulator equal to a problem it did not face a decade ago.