TO: The Hon. Evan Solomon, Minister of Artificial Intelligence and Digital Innovation;
The Hon. Marc Miller, Minister of Canadian Identity and Culture

FROM: Taylor Owen, Beaverbrook Chair, Media, Ethics and Communication, Founding Director, Centre for Media, Technology & Democracy, McGill University; Helen Hayes, Associate Director, Policy, Centre for Media, Technology & Democracy

DATE: February 24, 2026

RE: Scoping AI Chatbots into a revised Online Harms Act: The Case for Immediate Action

The Problem

Last week we learned that OpenAI's automated systems flagged the Tumbler Ridge shooter's ChatGPT conversations multiple times in June 2025, months before the attack that killed eight people. Internal employees debated whether to refer the case to Canadian law enforcement. Company leadership decided that the interactions did not meet their internal threshold but banned the account. Police were not contacted.

This tragedy has become another example of real-world harms caused by AI systems, including the Grok "undressing" crisis in January, where xAI's chatbot was used to generate millions of nonconsensual sexualized images of women and children, prompting regulatory investigations across Europe and Asia but exposing the absence of any Canadian enforcement mechanism.

In both cases design and safety decisions were made unilaterally by American companies, with direct consequences for Canadian users, and no Canadian regulator had knowledge of these protocols, nor the authority to scrutinize those decisions, ensure adherence to best practices, require mitigation measures, or enforce accountability before or after the fact.

The response to these failures cannot simply be to require companies to monitor and report private conversations to law enforcement. That raises its own serious concerns, particularly around privacy and the Charter. What is needed is a broader regulatory framework that addresses the upstream design decisions and safety architectures that allowed these situations to arise in the first place.

AI and Canada's Online Safety Policy

Online harms and AI consumer safety should no longer be considered separate policy portfolios, and the harms themselves should not be misunderstood as merely digital. The consequences of inadequate chatbot safety protocols are felt in schools and in communities. Consumer-facing AI chatbots present many of the same risks as social media platforms (including exposure to harmful content, risks to children, and potential facilitation of violence) but operate with even less transparency and fall outside of any existing Canadian regulatory framework. Continuing to treat them as distinct files, split across Heritage and AI portfolios with different legislative

timelines, leaves a widening gap that companies are already falling through, and deepens a tension between the government's stated safety and commercialization goals.

An updated version of the Online Harms Act, which should be treated as an immediate legislative priority, should include chatbots in its scope. The evidence base is already substantial and growing.

What We Would Already Know

Minister Solomon summoned senior OpenAI executives to Ottawa to explain their safety protocols. That was the right instinct, but it should not have been necessary. Had Canada established an online safety regulator that included chatbots in its scope, the government would already know how these companies flag dangerous content, what their escalation thresholds are, how they handle cross-border referrals, and whether their systems are adequate. We would not be learning about OpenAI's internal safety failures from the Wall Street Journal. We would know, because they would be required to tell us.

Applying the Part 1 of the proposed online harms act framework to chatbots would require, among other obligations:

Transparency over safety protocols. How is content flagged? By whom? What are the thresholds for escalation to law enforcement, and how are cross-border referrals handled? Right now, these decisions are entirely internal and opaque.

Mandatory risk assessments. Companies should be required to identify and assess risks their products pose to users, including risks related to violence, self-harm, and exploitation of minors.

Risk mitigation plans. Published, reviewable plans for how identified risks are being addressed, not just internal policies that surface only after tragedy.

Age-appropriate design. Enforceable standards for chatbot interactions involving minors, covering both content safeguards and design features.

Enforcement and accountability. Punitive mechanisms for non-compliance, administered by an independent regulator rather than relying on voluntary corporate commitments.

This Is Now a Justified Action

It is understandable that the government has been hesitant to impose new regulatory obligations on American technology companies in the current trade environment. But Tumbler Ridge makes clear that, at its core, technology policy should not be a trade matter. Rather, it is a fundamental public and consumer safety matter and an issue of domestic sovereignty. Canada is not singling out American firms: it is applying baseline safety standards to products deployed in its jurisdiction, exactly as it does with pharmaceuticals, vehicles, food, and financial services.

Canada is now the clear laggard among its peers. The UK's Online Safety Act has been in force since 2025, with Ofcom actively investigating and fining platforms, including opening a formal investigation into X over the Grok deepfake crisis. Every EU member of the G7 is covered by the Digital Services Act, under which the European Commission has ordered document retention from X, opened formal compliance proceedings, and French prosecutors have raided X's Paris offices. Even Japan, which has historically favoured softer regulatory approaches, enacted new platform transparency and content removal obligations last year. In the United States, where federal platform regulation has stalled, states are moving. California's Companion Chatbots Act, which took effect in January, requires AI companies to implement safety protocols, age verification, and crisis referrals for minors, with a private right of action for families. Texas has prohibited the deployment of AI systems that encourage self-harm or violence. Canada is the only G7 country with no digital safety regulator or no online harms legislation.

Reporting, Privacy, and the Limits of a Law Enforcement Approach

The instinct will be to focus narrowly on mandatory reporting to law enforcement. But a reporting-first approach raises fundamental privacy questions that need to be addressed directly. Mandatory reporting of flagged conversations raises real privacy concerns: it would require companies to surveil and disclose private user interactions at a scale that sits uneasily with Charter protections. It also places the entire burden on a last-resort mechanism, one that only activates after harm is already imminent, rather than on the upstream design decisions, risk assessments, and safety architectures that could prevent these situations from arising.

OpenAI's own stated rationale for not reporting was that "over-enforcement" could harm young people if, for example, police showed up unannounced at their door. That concern is not entirely unreasonable. But it cannot be a company's unilateral call. A regulatory framework would establish the conditions, thresholds, and oversight under which those trade-offs are made transparently and with democratic accountability, rather than behind closed doors in San Francisco.

Recommendation

The government should table an updated version of parts 1 and 4 of the Online Harms Act that explicitly scopes in consumer-facing AI chatbots, as recommended in our submission to the AI Strategy Task Force, and should treat this as a priority for the current session. The policy architecture from the previous bill (the Digital Safety Commission, the duty to act responsibly, the duty to protect children, the risk assessment and mitigation framework) provides a ready foundation. What this requires is including chatbots in the definition of regulated services and moving this legislation forward swiftly rather than sequencing it behind separate AI strategy or privacy timelines.

As long as Canada lacks a digital safety regulator with jurisdiction over these products, the government will continue to be in a reactive posture, dependent on the internal safety judgments of foreign companies, judgments that have repeatedly proven inadequate, and on media reporting

to learn when those judgments fail. This government has positioned Canada as a leader among middle powers on AI governance. That ambition is well placed, but it needs a domestic foundation. Tabling a comprehensive revised Online Harms Act that includes chatbots would protect Canadians and strengthen Canada's standing to shape the standards that other middle powers are actively seeking.