Enhancing trust in AI through increased transparency

Consultation on AI transparency

Submitted 23 September 2026

Consultation Background

On June 4, 2026, Prime Minister Mark Carney launched AI for All, Canada's new national AI strategy. Over the next five years, this strategy will introduce new legislation, investments, and programs that ensure AI is adopted responsibly, in a way that truly serves all Canadians – building trust, expanding opportunities, and reinforcing control of our sovereignty.

At its core, AI for All is grounded in the principle that the benefits of AI must be shared widely, ensuring that workers, businesses, and communities across Canada can access and benefit from AI. To achieve these goals, the Government of Canada is opening a dialogue on opportunities to enhance AI transparency so Canadians and Canadian businesses can make informed decisions about using, adopting, and engaging with AI, and to ensure that governments, researchers, and civil society organizations have the information they need to shape AI policy and research.

This consultation invites you to share your views on the AI transparency issues that are most important to you, and your views on Canada's next steps. Read the discussion paper here: Enhancing trust in artificial intelligence through increased transparency.

All questions in the consultation are optional. Respond only to those that align with your interests and/or expertise. Responses, aggregate data, and selected quotes from this consultation may be posted publicly online. Please do not include any personal or confidential information.

We appreciate your important contribution to this consultation process. If you encounter any difficulties with the consultation or require accommodation, please contact AIConsultations-ConsultationsIA@ised-isde.gc.ca .

⚠ Important: To protect your privacy and anonymity, please do not share any personal information in your responses.

1) AI-Generated Content

1. **Would it help you trust what you see online if you could tell whether something was created by AI or by a person? When and for what kinds of content (e.g., image, video, audio) is it most important to know if something was created or modified by AI?**

Yes. Being able to distinguish AI-generated or materially AI-modified content from human-created content helps Canadians make informed judgments about authenticity, authorship, and credibility of AI sources, and more informed to make decisions on which to trust.

Our latest research on AI-generated climate disinformation illustrates how AI can enable tailored content that seems credible in order to manipulate and influence efforts targeting municipal decision-makers.

Disclosure should not imply that human-created content is necessarily reliable, or that AI-generated content is necessarily false or fake. Its purpose is to allow people to assess context, seek verification, and make informed decisions on trust.

Disclosure is most important where AI-created (i.e. synthetic) or altered content could reasonably be mistaken for real evidence, genuine expertise, or human originality – and where that mistake could cause real material harm.

Priority contexts should include:

- **Political, civic, and electoral information**, including candidate or public-official impersonations, political advertising, advocacy communications, election-related audio or video, and purported evidence of public events. AI-generated (ie. synthetic) content can scale dangerous mis/disinformation, political manipulation, and fraudulent impersonation in an already fragmented and politically-charged information environment – leading to serious consequences for democratic institutions and public trust.
- **News, public-interest information, and emergency communications**, especially material presented as factual reporting, documentary evidence, eyewitness footage, public health guidance, and official government communication.
- **Audio, image, and video content that depicts real people, places, or events**, particularly deepfakes, voice clones, and non-consensual intimate imagery. These dangerous formats carry a heightened risk because they can be perceived as direct evidence, and can be used for nefarious, harmful purposes.
- **Fraud-sensitive communications**, including material involving finances (eg banking payment instructions, charity donations), social services (eg identity verification), employment (eg recruitment), legal aid (eg legal services), healthcare, or urgent requests apparently made by family members or authority

figures. AI-generated synthetic voice, video, and imagery has already been shown to enable identity theft and fraud against Canadians.

- **Education, scientific, professional, and consumer information** where the provenance, reliability, source, or degree of human review is relevant to a user's decision.
- **AI modification of authentic content** where the modification changes the reasonable interpretation of the material. The duty to disclose should not be limited to content generated from scratch. It should also cover material AI-altered in ways that affect identity, speech, actions, location, timing, evidentiary meaning, or apparent authenticity.
- **AI-generated text that informs the public on matters of public interest**, particularly where no identifiable human editor has exercised meaningful editing control.

A proportionate framework should distinguish among:

Content context	Recommended approach
Low-stakes creative, fictional, satirical, or entertainment content	Platform-appropriate disclosure of AI-manipulated or generated content inspired by the EU Code of Practice on AI labelling: https://ec.europa.eu/commission/presscorner/detail/en/ip_26_1328
Personal or private AI-assisted content	Default data provenance access (where technically feasible), as well as visible disclosure where AI-assisted content likely to be mistaken for authentic human content
Commercial, professional, or public interest content	Clear and accessible visible disclosure, paired with data provenance access
Deceptive deepfakes, impersonation, fraud, or election-related deception	Prominent disclosure of content as high risk; rapid detection and response duties on oversight bodies and platforms; accessible reporting and redress mechanisms; and stronger enforcement consequences

Our latest public consultation strongly supports this position: more than 4 in 5 respondents (84.7%) of respondents wanted disclosure for AI-generated or altered images, video, or audio; 86.8% wanted disclosure in news, political, health, emergency, and public-safety information; and 92.4% supported clear standards for labelling AI-generated content.

2. **What would best help Canadians determine when content is AI-generated (e.g., invisible or visible watermarks, disclaimers, provenance metadata)?**

   (2000 characters limit)

No single mechanism is adequate. Evidence for Democracy recommends a **layered transparency architecture**:

1. **Visible, standardized labels at the point of exposure**
   - Use short, plain-language labels (e.g., “Created with AI,”) that are prominent before or during consumption, not hidden.
   - Ensure labels are accessible across languages, screen readers, mobile interfaces, and lower-bandwidth environments.
2. **Machine-readable provenance and cryptographically signed metadata**
   - Require interoperable, machine-readable provenance for AI-generated or modified content, disclosing: AI use, generation/modification status, tool/provider, timestamp, and edit chain where available.
   - Keep provenance technology-neutral and avoid dependence on a single proprietary vendor or standard.
3. **Robust watermarking where it can meet published performance thresholds**
   - Use invisible/covert watermarks for automated detection and platform-level moderation, but do not treat them as a public-facing substitute for clear visual disclosures.
   - Establish independent government testing for robustness, false positives/negatives, removal susceptibility, spoofing resistance, and cross-platform persistence.
4. **Platform-level context and verification tools**
   - Platforms should preserve provenance signals, display standardized labels, and provide easy “learn more” functions that explain what is known, what is unknown, and why a label was applied.
   - Absence of provenance should be labelled as “no verifiable origin information,”.

Technical measures are complementary rather than interchangeable. Provenance tracking, watermarking, metadata, and detection tools can help identify a content item’s origin and history, but none provides a comprehensive solution on its own. Their effectiveness depends on deployment, oversight, usability, and complementary regulatory, educational, and market measures.

3. Who in the AI value chain (developers, deployers, or others) should be responsible for providing transparency around AI-generated content? Why?

Responsibility should be **shared across the value chain**, but differentiated according to each actor's control, knowledge, and capacity to prevent and mitigate harm. A governance approach focused only on developers would fail because disclosure can be lost, omitted, or made to mislead Canadians while AI systems are deployed, distributed or promoted across the value chain.

Actor	Core responsibilities
Foundation-model and generative-AI developers	Build interoperable provenance and watermarking capabilities into systems; document their technical limitations; publish detection tools or interfaces where appropriate; conduct independent testing.
Product developers and service providers	Preserve and expose data provenance in all their products; provide clear default disclosure features; enable downstream users to make accurate disclosures
Deployers and organizational users	Give visible, context-specific disclosure when publishing or using AI-generated/synthetic content in professional, public-facing, or high impact settings; ensure a meaningful human reviewer is ultimately accountable when deployers or organizations claiming human review or editorial control.
Publishers, advertisers, broadcasters, and political actors	Verify data provenance where reasonably possible; label synthetic or materially altered content; maintain records; establish elevated procedures for public-interest and election-related material.
Online platforms and intermediaries	Preserve provenance metadata; display interoperable labels; provide user reporting and appeal systems; limit algorithmic amplification of undisclosed deceptive synthetic content; share systems data with researchers and regulators.
Public institutions and government contractors	Establish higher standards for disclosure and governance, including prominent notices, publication of data provenance information, clear procurement requirements, human accountability mechanisms, and records management.
Regulators and standards bodies	Set baseline requirements; accredit and recognize standards; support testing; monitor implementation; enforce duties; and update rules as technologies and risks evolve.

Developers must not be able to continue to externalize responsibility simply because a downstream actor chose the prompt or publication channel – as seen in the case of OpenAI's role in enabling the Tumbler Ridge massacre.

Conversely, deployers and publishers are generally best placed to assess the specific communication context, audience, and risk of deception. Best practices can be drawn from the **EU AI Act's Article 50 model,** which separates provider duties (such as machine-readable marking) from deployer duties to disclose deepfakes and certain text made in the public interest.

4. **Do existing market practices, technical tools, and legal frameworks make it easy enough to know when content is AI-generated? If not, where do gaps remain and what actions (e.g., regulatory measures, guidance and codes of conduct, standards and technical solutions, research and development, literacy initiatives, procurement requirements) do you think the Government should take?**

No, existing market practices, voluntary commitments, technical tools, and general legal frameworks are not sufficient to ensure that Canadians can reliably identify AI-generated or AI-modified content.

ISED's earlier voluntary code of conduct on AI usage encouraged developers to provide a reliable, freely-available means to detect AI-generated content (with particular short-term attention to audio-visual content), and suggested that organizations identify systems that could be mistaken for humans. These are useful starting points, but voluntary measures have predictable shortcomings: incomplete coverage, inconsistent implementation, limited interoperability, opaque performance claims, and few to no consequences for non-compliance.

Key gaps include:

- No broadly enforceable Canadian baseline duty to mark AI-generated or materially altered content.
- Fragmented, proprietary, and inconsistent technical implementations across industry.
- Weak preservation of provenance when content moves between services or is downloaded, re-uploaded, cropped, compressed, or transformed.
- Insufficient independent testing of watermarking and AI detection systems.
- Significant challenges for text, multilingual content, live generation, open-weight models, and content made through tools outside major commercial platforms.

- Lack of common user-interface rules for labels, explanations, accessibility, and consent.
- Insufficient obligations on platforms to preserve, display, and disclose data provenance.
- Insufficient transparency about automated moderation and enforcement decisions.
- Unequal impacts on communities targeted by impersonation, non-consensual imagery, fraud, and political disinformation.
- Absence of sustained access for public-interest researchers to access data provenance of platforms, in order to determine whether labeling and provenance measures work in practice.

Evidence for Democracy recommends that the Government of Canada:

- **Create enforceable, risk-based transparency obligations** for AI-generated and materially AI-modified content, beginning with high risk and public-facing contexts.
- **Adopt a layered disclosure requirement** to include visible disclosure for people, interoperable machine-readable provenance for systems, and robust watermarking where independently validated and appropriate.
- **Define "material modification" clearly**, focusing on changes that affect the interpretation of AI-generated vs. genuine human identity, speech, and content. Ordinary editing, accessibility adaptations, spell-checking, compression, and minor aesthetic changes should not trigger disproportionate duties.
- **Establish common standards and testing protocols**, ideally through international standards bodies (e.g. International Organization for Standardization) and Canadian participation in those processes. Interoperability, accessibility, privacy, security, and open verification should be required design objectives.
- **Create an independent evaluation and assurance regime** for technical provenance, watermarking, and detection tools. Vendors should be required to publish performance data by modality, language, content type, and transformation scenario, including false-positive and false-negative rates.
- **Set platform duties** to ensure data provenance access, display standardized disclosures, provide reliable reporting and redress, and assess systemic risks associated with the reach and amplification of undisclosed AI-generated/synthetic content.
- **Apply special rules to democratic and public interest information**, including election communications, political advertising, official government communications, journalistic content, and material likely to affect public safety or access to services.
- **Use federal procurement as a market-shaping tool.** Government should require suppliers of AI systems and digital content services to meet interoperability, disclosure, auditability, accessibility, privacy, and incident-reporting requirements.

- **Fund independent research and civil society access** to evaluate policy effectiveness and organizational compliance, impacts on vulnerable communities, and the information integrity effects of AI disclosure regimes.
- **Invest in digital media literacy**, while avoiding short-sighted approaches that shift all responsibility onto individuals (e.g. parents) or platform users (e.g. youth). Holistic digital media literacy should help Canadians interpret labels and critically-assess sources; it cannot replace duties on powerful hyperscalers, developers, deployers, and social media platforms.
- **Coordinate with privacy, consumer protection, competition, election, and human-rights regulators**, since synthetic content risks often cross institutional mandates.

2) AI Interaction

1. **When and why is it most important to know that you are interacting with an AI system? Are there contexts where you don't need to know that you're interacting with AI?**

People should be told when they are interacting directly with an AI system that does (or could reasonably) affect their choices, rights, dignity, privacy, safety, or ability to seek human assistance.

It is especially important in the following contexts:

- **Government and public services**, including immigration, benefits, taxation, housing, policing, corrections, social services, and emergency or health-related communications.
- **High-impact private services**, such as employment, education admissions, credit, insurance, housing, healthcare, legal services, financial services, and essential utilities.
- **Consumer transactions and customer service**, particularly where an AI system recommends products, resolves disputes, collects personal information, makes representations about rights or eligibility, or affects access to a human representative.
- **Children, youth, and other vulnerable users**, including in educational, mental-health, caregiving, or persuasive contexts.
- **Systems that can be mistaken for a person**, including chatbots, voice assistants, avatars, simulated agents, and synthetic customer-service representatives.
- **Systems that collect or infer sensitive information**, including emotion-recognition, biometric-categorization, behavioural-profiling, and

persuasive systems.
- **Contexts involving advice, authority, or dependency**, where users may over-rely on the system or believe it has professional credentials or credibility, legal authority, empathy, confidentiality, or accountability that it does not possess.

The rationale is not simply that users have a right to know the technical mechanism being used. It is that disclosure supports informed consent, helps people calibrate trust and reliance, makes it possible to challenge or escalate a decision, and reduces deception and manipulation. The **EU AI Act** has categorized **levels of AI risk**, outright banning "unacceptable AI" (e.g. financial manipulation and election impersonation technologies), while requiring strict measures to inform users when they are interacting with "high risk" AI systems.

There are limited contexts where an additional, repeated disclosure may not be necessary:

- When the AI nature of the interaction is already **obvious** to a reasonably-informed user from the context and interface (for example, a clearly labelled "AI image generator" or an embedded grammar tool).
- When an individual has already received a **clear disclosure** in an ongoing interaction and can readily access a persistent reminder.

These should be narrow exceptions. "Obviousness" should not become a loophole that permits organizations to hide AI use through human-sounding names, realistic avatars, conversational design, or buried disclosures.

2. What factors make a disclosure of AI use meaningful and effective for the user?

A meaningful disclosure must be **timely, clear, specific, accessible, and actionable**. Merely mentioning that "AI may be used" in a Terms of Service agreement or placing a faint icon in a corner of an image is not meaningful transparency.

Effective disclosures should:

- Appear **before or at the start of the interaction**, rather than after the user has relied on the system or shared sensitive information.
- State plainly that the user is interacting with AI (e.g. "You are chatting with an AI assistant, not a human agent.")
- Identify the **organization responsible** for the interaction and provide a means to contact them.
- Explain the AI system's relevant purpose, capabilities, and material limitations in context. A health navigation chatbot, for example, should state that it is not a

clinician, describe what it can assist with, and explain how to obtain human help from a medical professional.

- Identify if and when the user's content will be recorded, used to improve a system, shared with third parties, or used for profiling or personalization
- Provide an easily accessible route to **a qualified human**, complaint process, correction mechanism, or alternative service channel when the context is high impact or the user requests it.
- Be presented in accessible formats and in Canada's official languages, with attention to plain language, disability accessibility, low digital literacy levels, and diverse communication needs of Canadians.

A useful distinction is between disclosure of **identity** and disclosure of **function**:

Disclosure element	Example
Identity	"You are interacting with an AI system."
Purpose	"This AI tool helps summarize your application and identify missing documents."
Limits	"This AI chatbot may make mistakes and cannot determine eligibility."
Consequences	"A human officer makes the final decision."
Recourse	"To speak with a person or challenge information, select 'Speak with Human Agent' "

3. **Who in the AI value chain (developers, deployers, or others) should be responsible for providing transparency around AI interactions? Why?**

As with content transparency, responsibility should be shared but differentiated across the value chain.

- **Developers** should make systems technically capable of clear disclosures, record-keeping, complaint and escalation processes, and safe handoff to humans. They should document intended uses, foreseeable misuse, limitations, and any human-like features that may create confusion.
- **Deployers** should bear the primary responsibility for user-facing disclosures because they control the interface that users use, and service design (e.g. a bank,

university, employer, government department, insurer, or platform cannot avoid responsibility by pointing Canadians to a third-party model provider).

- **Organizations that procure or integrate AI** should conduct pre-deployment impact assessments, test disclosures with affected users, ensure pathways for users to escalate issues, and maintain accountability for outcomes.
- **Platforms and app stores** should be held to a higher standard given the power imbalance that exists between Canadian users and megacorporations, and be required to ensure that AI agents and conversational systems distributed through their services meet baseline labelling, accessibility, and reporting requirements.
- **Public institutions** should be held to a higher standard given their role in shaping AI governance standards, guidelines, and policy decisions.
- **Regulators** should define minimum requirements, provide guidance, investigate complaints, audit systemic practices, and coordinate enforcement across privacy, consumer protection, human rights, and sector-specific structures.

4. **Are existing market practices and legal frameworks sufficient to support transparency around AI interactions? If not, where do gaps remain and what actions (e.g., regulatory measures, guidance and codes of conduct, standards and technical solutions, research and development, literacy initiatives, procurement requirements) do you think the Government should take?**

No, existing practices are inconsistent, often lacking transparency, and typically governed by voluntary design choices rather than enforceable user rights.

Many existing systems make AI disclosure easy to miss, use ambiguous language, or offer no clear policy path to ensuring transparency around AI usage. General privacy and consumer-protection rules may address some deceptive practices (e.g. Bill C-16, the Protecting Victims Act), but they do not provide a comprehensive, AI-specific framework for disclosure, explainability, human recourse, record-keeping, or accountability.

Canada's voluntary generative-AI code of conduct previously recognized that systems that might be mistaken for humans should be clearly and prominently identified as AI systems. This principle should now be made enforceable, and supported by implementation guidance and oversight.

We recommend that the Government of Canada:

- **Establish a baseline statutory duty** to disclose direct AI interaction clearly and at

the first meaningful point of contact, subject only to narrow, context-sensitive exceptions.

- **Require heightened safeguards in high-impact and public sector contexts**, including meaningful human review, accessible escalation pathways to human oversight, documentation, record retention, and a clear mechanism to contest or correct consequential outcomes.
- **Prohibit deceptive interface practices** that obscure AI identity, falsely imply human review, overstate capability, or create misleading impressions of authority, confidentiality, emotion, or professional expertise.
- **Develop model disclosure templates and usability guidance**, with testing requirements for plain language, official languages, and accessibility for diverse user groups.
- **Require organizations to publish meaningful and accessible information about deployed AI systems**, including purpose, capabilities, limitations, intended users, human oversight structures, data practices, and pathways for recourse.
- **Require impact assessment and independent audit** for high-impact systems, with public summaries and meaningful consultation with affected communities where systems are deployed at scale.
- **Create incident-reporting, complaint, and redress pathways** that allow individuals and civil society organizations to report misleading AI disclosures and harmful practices, lack of transparency measures, denial of human access to data provenance, and other systemic failures.
- **Use procurement rules to ensure government leadership.** Federal institutions should not deploy public-facing AI without clear disclosure, human alternatives, records of system use, accessibility testing, privacy safeguards, and independent evaluation.
- **Support public-interest research and regulator capacity.** Civil society, academic researchers, journalists, and regulators need lawful, privacy-protective access to information needed to evaluate AI system deployment patterns, algorithms, disclosure effectiveness, discrimination measures, and other systemic impacts.
- **Coordinate a coherent framework across federal authorities.** Transparency around AI interactions implicates the Privacy Act or any successor legislation to PIPEDA (e.g. Bill C-36 - the Protecting Privacy and Consumer Data Act); consumer protections, accessibility obligations, or election integrity structures.

Cross-cutting principles

We encourage the Government to adopt the following principles across both content and interaction transparency:

- **Proportionality:** Obligations should increase with the likely severity, scale, and irreversibility of harm.
- **Meaningful human control:** Disclosure should support (not replace) real human accountability, ability to contest decisions, and access to recourse mechanisms.
- **Interoperability:** Canadians should not need a different verification tool for every platform or vendor.
- **Privacy and security by design:** Provenance systems should minimize unnecessary collection of Canadians' personal data, and avoid exposing sensitive location or identity information.
- **Accessibility and inclusion:** Transparency tools must work for people using assistive technology, for users with different levels of digital media literacy, and across Canada's linguistic, ethnic, and regional diversity.
- **Independent evaluation:** Claims about watermarking, content detection, and disclosure effectiveness should be tested by independent experts against realistic cross-platform conditions.
- **Public-interest accountability:** Researchers, journalists, civil society, and regulators need sufficient access to AI systems in order to assess whether transparency measures are functioning, rather than relying solely on faulty industry self-reporting.
- **No false assurance:** A label, watermark, provenance record, or absence of such a signal must never be represented as definitive proof that content is true, false, human-created, or harmless. Technical provenance can support credibility assessment, but it cannot by itself establish truthfulness or contextual accuracy.

3) Information about AI Systems

1. **What kinds of information about AI would you like to know in order to help you make informed choices about when and how to use AI? When and for which types of AI products is this information most important to you?**

To make informed choices about when and how to use AI, researchers and evidence-based organizations like Evidence for Democracy need access to **consistent, comprehensible, useful, and reliable information** about AI systems across several dimensions:

- **Purpose and intended use cases**: Clear statements of what the AI system is designed to do, including any known harmful or inappropriate uses.
- **Training data provenance**: High-level descriptions of the categories of data used to train the model, including whether personal data, copyrighted works, or publicly-scraped content were involved.
- **Known limitations and risks**: Documentation of contexts in which the AI system should not be used, and any known biases or performance disparities across demographic groups.
- **Human oversight mechanisms**: Information about whether (and if so, how) human review and oversight is integrated into the system's operation, particularly for consequential decisions on sensitive areas (eg immigration, social services).

This information is **most critical** in the following contexts:

- **High-stakes or consequential decisions** (i.e. high risk uses of AI): When AI is used to screen job applicants, assess creditworthiness, determine eligibility for public benefits, or inform health-related guidance.
- **Public-facing or persuasive content**: When AI generates text, audio, or visual content intended to inform the public on matters of public interest, including news, policy analysis, or educational materials.
- **Research and evaluation**: When researchers need to audit, replicate, or assess the reliability of AI-generated findings or tools used in evidence synthesis.

2. Who in the AI value chain (developers, deployers, or others) is best placed to provide what information about AI systems, how they are created, and their capabilities and limitations?

Responsibility for transparency should be **distributed across the AI value chain**, with obligations calibrated to each actor's control and expertise:

Actor	Primary Disclosure Responsibilities	Rationale
Developers (providers)	• Training data categories • Documenting known limitations, intended and harmful uses,risks • Technically marking AI-generated/synthetic outputs (e.g., watermarks, metadata) • Updating and versioning notices	Developers have the deepest knowledge of system design, training, and capabilities. They are best placed to provide foundational technical documentation and embed detectable markers in outputs.
Deployers	• Context-specific notices to end-users (e.g., "You are interacting with an AI system") • Disclosure of AI usage in consequential decisions (e.g., credit decisions) • Labelling of deepfakes and AI-generated text • Incident reporting for serious harms	Deployers control the context of AI use, and are best positioned to inform affected individuals about how an AI system impacts them. They also bear responsibility for ensuring that provider disclosures are communicated effectively to end-users.
Distributors & integrators	• Forwarding developer disclosures to deployers • Ensuring compatibility of transparency mechanisms (e.g., metadata standards)	Intermediaries play a critical role in maintaining the integrity of disclosure chains, particularly in complex supply chains.
Regulators & standards bodies	• Maintaining public registries of certified AI systems • Publishing aggregated incident data • Developing interoperable technical standards for marking and detection	Public institutions can provide the infrastructure for accountability – better enabling researchers and civil society to monitor systemic risks.

This division of labour aligns with the EU AI Act's Article 50, which assigns technical

marking obligations to providers, and human-facing disclosures to deployers.

3. **Are existing disclosure practices and legal frameworks sufficient to provide the information that users and other stakeholders need? If not, where do gaps remain and what actions (e.g., regulatory measures, guidance and codes of conduct, standards and technical solutions, research and development, literacy initiatives, procurement requirements) do you think the Government should take?**

Existing frameworks are insufficient. While Canada's Voluntary Code of Conduct, provincial measures (e.g., Ontario's AI hiring disclosure requirement), and proposed federal legislation (e.g., Bill C-34, Bill C-36) represent important steps, significant gaps remain for Canada:

- **No mandatory disclosure regime**: Unlike the EU, Canada does not yet require developers or deployers to disclose AI-generated content or AI interactions in most contexts.
- **Fragmented and inconsistent practices**: Market-led voluntary disclosures (eg ISED's Code of Conduct on AI usage) vary widely in clarity, granularity, and enforceability.
- **Limited recourse for affected individuals**: Canadians lack clear mechanisms to contest AI-driven decisions or verify the authenticity of AI-generated content.

Recommended actions:

1. **Enact targeted transparency obligations**: Introduce legislation requiring developers to mark synthetic outputs and deployers to disclose AI interactions in high-risk contexts, modelled on Article 50 of the EU AI Act.
2. **Develop standardized disclosure templates**: Issue guidance or regulations specifying the minimum information developers must provide to deployers and end-users, including intended uses, training data categories, and known limitations.
3. **Establish an AI incident reporting mechanism**: Create a centralized, public-facing database for serious AI-related incidents, mirrored on product safety reporting systems.
4. **Invest in digital media literacy and public education**: Fund digital media literacy organizations and initiatives to help Canadians interpret transparency signals (e.g., labels, watermarks) and understand AI's role in decision-making.
5. **Leverage procurement and certification**: Require government suppliers to meet transparency standards and participate in the *Trusted AI Certification* program.
6. **Support research and technical standards**: Fund R&D into robust, interoperable

marking and detection technologies, particularly for text and multimodal content.

By adopting a **layered, risk-proportionate approach**, the Government can enhance transparency without imposing undue burdens on innovation or exposing legitimate trade secrets.

4) AI Incidents

1. **Do the public and the Government have enough information about AI risks, incidents, and potential future harms? What types of information related to AI use and risk trends would be most useful so that Canadians and the Government can proactively address AI-related risks and harms?**

No, the Canadian public lacks sufficiently accessible information about AI risks, incidents, near misses, and emerging harms. Instead, information is unevenly dispersed across providers, deployers, regulators, researchers, media, civil society, and affected people. This prevents early detection of recurring failures and leaves the public and regulators dependent on voluntary disclosure or after-the-fact controversy (eg OpenAI and Tumbler Ridge)

Canada should establish a coordinated national AI incident and risk-information system that connects sectoral regulators and is compatible with the OECD AI Incidents Monitor and common reporting framework.

Useful information includes:

System identity, provider, model or version, purpose, sector, and deployment context.
Whether AI generated, recommended, ranked, decided, communicated, or took external action.
Data sources, privacy practices, sensitive-data use, and cross-border data flows.
The extent of human review, overrides, appeals, complaints, corrections, and errors.
Affected people and the type, scale, duration, and distribution of harm.
Incidents, near misses, vulnerabilities, material model updates, and misuse patterns.
Evidence of bias, privacy breaches, cybersecurity failures, false outputs, fabricated citations, environmental impacts, and misinformation or impersonation risks.
Mitigation measures, notifications, remediation, suspension, recall, and residual

risk.

For public-sector use, disclosure should include what an AI system did, what evidence informed its outputs, whether sources were independently verified, who reviewed the work, and which official remains accountable. Our AI and Democracy research report identifies risks from inaccurate AI-generated research, fake citations, and undisclosed "shadow AI" use in government (eg Government of Newfoundland and Labrador's redacted education action plan)

This aligns with the consultation: respondents wanted more than a label—they wanted to know how AI was used, what informed it, how much human review occurred, and who remained responsible.

2. **How should a serious AI incident be defined, and what information would be most useful to better understand the causes and consequences, as well as the mitigation, of these incidents? How could reporting requirements be structured to encourage proactive disclosure of AI incidents and also address concerns around confidentiality and liability?**

A **serious AI incident** should mean an event, condition, or series of related events in which an AI system's development, deployment, use, malfunction, or misuse causes (or creates a substantial likelihood of causing) serious harm.

This definition should include:

- serious physical, mental-health, financial, privacy, discriminatory, environmental, cybersecurity, public-safety, or rights-related harm.
- wrongful restrictions on access to essential services or benefits;
- disruption of critical infrastructure or government services; and
- AI-enabled impersonation, fabricated evidence, misinformation, or coordinated influence that materially threatens public health, democratic participation, or trust in institutions.

Canada should require reporting of significant near misses and systemic hazards, not solely realized harm. For example, a discriminatory AI-generated recommendation caught by an immigration agent, or fabricated AI citations detected before publication may reveal a recurring failure that could cause harm at scale.

Reports should document:

The system, provider, deployer, model/version, purpose, and connected tools.

What occurred, who was affected, the severity, and whether harm was actual, imminent, or avoided.
The system's role and the human-review process.
Relevant data, prompts, logs, evidence sources, and technical configuration.
The cause: model behaviour, data, integration, security, human error, automation bias, governance failure, or misuse.
Immediate containment, notification, remediation, compensation, system changes, and residual risk.

In terms of reporting concerns, we recommend the government use staged reporting: a quick initial notification, followed by an interim mitigation report, a final root-cause and remediation report, and an anonymized public summary. Reporting should begin on a reasonable likelihood of AI contribution, not definitive proof. Inspiration can be drawn from the EU AI Act, which outlines a relevant severity-based reporting model.

Ultimately, good-faith early reporting should not be treated as an admission of liability, while non-reporting, concealment, recklessness, or failure to mitigate should carry consequences.

3. **Which individuals or actors (e.g., developers, deployers) are best placed to provide the most relevant information? Who (e.g., regulatory bodies, nonprofits, others) should collect this information and what should be done with it?**

Responsibility should be shared across the AI lifecycle, with duties reflecting each actor's knowledge, capacity, and control.

- **Providers and developers** should provide information on report model capabilities, limitations, testing, red-team results, data governance, updates, vulnerabilities, known failure modes, incident patterns, and remediation.
- **Deployers and operators** should provide reports on the actual deployment context, including: inputs, prompts, integrations, permissions, human oversight, affected people, complaints, appeals, adverse outcomes, near misses, and local corrective action.
- **Integrators, contractors, distributors, and cloud providers** should report where their configuration, hosting, modification, or distribution materially affects system performance or risks.
- **Public institutions** should place enhanced obligations on systems used in public services, administrative decisions, policy development, public communications, enforcement, and procurement. Vendor confidentiality should not prevent

disclosure where public rights, expenditures, services, or trust are affected.

- **Affected people, workers, researchers, journalists, civil society organizations, and whistleblowers** require safe, accessible channels to submit complaints and evidence in cases where harms are occurring.

Canada should create a coordinated model in which sectoral regulators retain their mandates, while a federal coordinating authority maintains common definitions, reporting templates, severity thresholds, public-reporting standards, and cross-sector analysis. An independent body should publish aggregate trends and identify systemic risks.

In our previous research, we have recommended a federal oversight body, the Digital Safety Commission, be empowered with enforcement powers, audit authority, information-gathering powers, and ability to issue penalties and require corrective action of violators.

Ultimately, information should serve the public interest, with the goal of supporting enforcement, strengthening public warnings, recalling or suspending violators, issuing remedies, and developing standards —not simply internal compliance reporting.

5) AI Agents

1. **Are you concerned about the use of AI agents? In what contexts is it most important to know when an AI agent is in use? Why?**

Yes, AI agents raise heightened concerns because they may do more than generate information: they can retrieve data, use tools, access systems, communicate externally, alter records, make purchases, submit forms, and initiate transactions.

Our principal issue with AI agents is that they represent delegated authority (away from human decision-makers) that uses Canadians' sensitive data (often without their knowledge) without clear accountability or any clear oversight mechanisms.

Disclosure is most important where an AI agent:

- **Interacts with the public in an official capacity** (e.g. as a government agency chatbot)
- **Handles highly sensitive data** (e.g. Canadians' personal, financial, health,

employment, educational, or biometric information).

- **Influences Canadians' fundamental rights and personal outcomes** (e.g. Constitutional rights, access to employee benefits, access to healthcare).
- **Operates in high-risk areas**, namely government services, emergency management, public communications, law enforcement, health, finance, or critical infrastructure.
- **Can act in an official capacity to make substantial decisions** (e.g. make purchases, initiate payments, send official communications, alter records, access sensitive databases).
- **Produces outputs that influence policy outcomes** (e.g. policy briefs, patient charting, police reports, immigration applications).
- **Interacts with vulnerable populations** (e.g. children, seniors) who may reasonably believe they are communicating with a human.
- **Is used to influence government decision-making or elections** (e.g. public consultations, municipal councils, elections, or other democratic processes).

We believe that Canadians should know not merely that "AI is involved" in decisions that affect them, but whether the agent has the authority to act, what information it uses, which organization is responsible, and how a human can intervene and/or address AI-related harms.

Our latest public consultation supports this approach: 89.6% of respondents said AI use should always be disclosed; 85.4% said they wanted disclosure where AI contributes to decisions affecting people; and 82.6% indicated they desired it where sensitive information is involved.

Broader considerations

1. **Given the areas discussed here, when is AI transparency most important, and are there any areas where action would be premature? Are there other areas that should be considered for action to enhance AI transparency?**

Transparency is most important when AI can materially affect Canadian individuals, communities, public institutions, constitutional rights, democratic processes, and the information ecosystem.

Priority contexts include systems obscuring access to fundamental services and freedoms; systems using sensitive data; systems that profile, rank, monitor, or decide about people; government AI; and systems that create policy-related outputs (eg

AI-generated policy briefs) or deal with public interest information (e.g. health biometric data).

Our consultation supports a **risk-based approach.** E4D community respondents were relatively comfortable with AI being used for generic administrative tasks and information organization (60.4%), but only a quarter (24.3%) were comfortable with AI decisions about programs or benefits, and less than 1 in 6 were favourable towards AI usage in immigration, tax, or enforcement decisions (13.9%), and virtually no one was comfortable with AI decisions made about Canadians without human oversight or review (3.5%).

Action, when it comes to AI, may be premature where a use is genuinely speculative or a rigid technology-specific rule would become obsolete. But uncertainty should not delay baseline obligations where systems are deployed, affect constitutional rights, operate at scale, or could cause irreversible harm to individuals and communities. Canada should regulate the function, authority, data sensitivity, and impact of AI — not simply add marketing labels on AI products.

Additional transparency priorities include:

- Environmental impacts, including on energy, water, emissions, and AI infrastructure.
- Copyright, consent, likeness, training data governance, and creator rights.
- Accessibility, language rights, and unequal access.
- Indigenous data sovereignty and community governance.
- Labour, workplace surveillance, professional operations, and automated hiring.
- Vendor dependence, procurement, data sovereignty, and cross-border data infrastructure.
- AI-generated policy materials and products
- Children's rights and protection of vulnerable and marginalized groups.

Our research on public trust shows that weak evidence practices and opaque decision-making can erode confidence in institutions. Disclosure must therefore include evidence quality, uncertainty, source verification, human review, and responsibility, and not only the fact that AI was used.