FROM TRANSPARENCY TO HUMAN CONTROL

A Systems-Engineering Response to Canada's Consultation on Artificial Intelligence Transparency

Submitted to: Innovation, Science and Economic Development Canada
In response to: Enhancing Trust in Artificial Intelligence Through Increased Transparency
Discussion Document, July 2026
Submitted by: Frank Misa — Canadian Software Architect and Systems Engineer
September 2026

Executive Summary

Canada's discussion document identifies five important areas of artificial-intelligence transparency: identifying AI-generated content, knowing when one is interacting with AI, understanding AI systems and their limitations, reporting serious AI incidents, and tracking the activities of AI agents.

These are appropriate areas of inquiry. My principal concern is that transparency should not become an end in itself.

Knowing that AI was involved is useful. Knowing what an AI system can and cannot do is useful. Recording that an AI system failed is useful. But meaningful accountability requires an additional question: What can a person actually do with that information when something goes wrong?

From a systems-engineering perspective, transparency should therefore be connected to operational controls: human recourse, independent testing, reliable audit records, identifiable system and agent identities, incident reconstruction, and the ability to intervene when consequential automated processes fail.

I suggest organizing these protections as a three-layer Rights Stack:

01 HUMAN AUTHORITY — Humans remain sovereign over consequential machines.

KNOW · READ · APPEAL · OVERRIDE · AUDIT · STOP · OFFLINE

02 HUMAN PARTICIPATION — Automation must benefit the people and societies that make it possible.

SHARE · OWN · FAIR APPLICATION

03 HUMAN INTEGRITY — Technology must not take ownership of the person.

PRIVACY · IDENTITY · CONSENT · REMAIN HUMAN

The present consultation most directly engages Human Authority. The other two layers identify broader economic and personal questions that increasingly capable AI will also raise. The framework does not reject automation; it is intended to keep automation answerable to human beings.

1. Transparency Should Lead to Human Recourse

The discussion document correctly recognizes that merely notifying someone that AI is present may not constitute meaningful transparency. It notes that useful disclosure may need to describe the role of an AI system, how personal information is used, and what human oversight mechanisms are in place. It also distinguishes casual AI interactions from consequential situations such as employment, health and financial

decisions. (Discussion Document, p. 15.)

I believe this points toward a broader principle: A notice saying that AI was used is not meaningful transparency if the affected person has no effective means of challenging what the AI did.

This concern extends beyond AI. Canadians already encounter automated customer-service systems, social-media platforms and digital workflows in which an account can be suspended, an application rejected or some other consequential action taken without an obvious route to an accountable human being. Artificial intelligence could significantly magnify this problem.

For consequential applications, disclosure should therefore be paired with meaningful human recourse. Someone materially affected by an automated process should be able to determine that AI materially participated; what role it played; who is legally or institutionally responsible for the resulting action; how the affected person can challenge the result; and whether an accountable human can review and, where justified, reverse it.

This need not require a telephone call for every trivial automated interaction. The requirement should be proportionate to the consequences. But an organization should not be permitted to use automation to create an accountability dead end.

2. Canada Should Independently Test Important AI Systems

The discussion document considers disclosure concerning AI capabilities, limitations, appropriate uses, development methods and training data. It also recognizes the practical difficulty of obtaining detailed information about proprietary systems and training datasets. (Discussion Document, pp. 18–20.)

There is another approach worth emphasizing: Do not rely exclusively on what an AI provider says about its system. Test how the deployed system actually behaves.

Canada already has institutions working on AI safety, certification and standards. The discussion document notes investments in the Canadian AI Safety Institute, a Trusted AI Certification program and the Standards Council of Canada’s AI work. (Discussion Document, p. 5.) These initiatives create an opportunity to develop reproducible Canadian testing methodologies.

For high-impact deployments, accredited testing could examine characteristics such as factual reliability within a defined domain, resistance to adversarial or prompt-injection attacks, inappropriate use of permissions, reliability of safeguards, consistency, uncertainty disclosure, human-override mechanisms and behaviour when tools or external systems are connected.

This would be analogous to conformity and safety testing in other technical fields. The objective should not be for government to establish an ideological or political ‘truth test’ for general-purpose AI. Instead, tests should have defined methodologies, reproducible conditions and measurable results tied to particular applications.

For example, testing whether an AI medical system fabricates citations under defined conditions is an engineering question. Testing whether a general-purpose AI expresses an officially approved opinion is not. Independent testing could provide useful information even where developers cannot or will not disclose proprietary details of model architecture or training data.

3. Provenance Must Be Visible at the Point of Use

The discussion document already addresses hidden watermarks, provenance metadata and technical

standards such as C2PA. It also recognizes their limitations, including removal or evasion by malicious actors and possible privacy implications. (Discussion Document, pp. 11–13.) I therefore do not suggest that Canada invent another incompatible metadata standard.

Instead, Canada could concentrate on making existing provenance mechanisms useful to ordinary people. There is an important distinction between provenance information existing inside a file and a person actually seeing and understanding that information when the content reaches them.

Canada could encourage or develop common interfaces that allow browsers, email clients, messaging applications, social platforms and other software to surface provenance information consistently. An ordinary user might see a simple indication that provenance information is available; content was materially generated or modified by AI; the source cannot be established; or expected provenance information is missing or inconsistent. Additional technical information could be available to anyone who chooses to inspect it.

This approach also recognizes an important practical reality: many AI systems and much AI-generated content encountered by Canadians will originate outside Canada. Canada may have limited ability to dictate how a foreign model is constructed, but it has considerably more ability to influence the standards and interfaces through which AI-generated material reaches Canadian consumers, businesses and institutions.

4. Require Audit Traces, Not Artificial “Thought Transcripts”

As AI systems become agentic, disclosure becomes substantially more complicated. A conventional chatbot produces an answer. An AI agent may instead access a database, read email, invoke an API, communicate with another agent, purchase something, transfer information or alter an external system.

The discussion document recognizes this problem. It describes emerging AI observability systems that record tool use, agent activity and decision pathways, as well as work on common telemetry standards and agent credentials. (Discussion Document, pp. 27–29.) I believe Canada should encourage the development of a minimum interoperable audit standard for consequential AI agents.

This should focus on what a system actually did rather than attempting to expose a model’s private internal ‘chain of thought.’ An audit trace for a high-impact agent could record the model and software version involved, the responsible operator, the authenticated identity of the agent, permissions granted to it, significant tool or API calls, external actions taken, interactions with other agents, applicable policy versions, human approvals and timestamps.

Where possible, this logging should occur outside the agent’s direct control. That is an important security principle. A system being audited should not possess unrestricted ability to erase or rewrite the evidence required to audit it. Tamper-evident logging, external execution gateways and independently maintained records may therefore become increasingly important as agents acquire greater autonomy.

I would not recommend requiring AI systems to expose private internal reasoning or purported ‘thoughts.’ Such representations may not reliably explain causation and can create additional privacy, intellectual-property and security problems. The appropriate goal is a decision and action trace: What information entered the consequential process? What system acted? What tools did it use? What authority did it possess? What did it change or transmit? What external actions resulted? Which humans approved or subsequently reviewed those actions? Those are questions that can support accountability.

5. AI Agents Should Have Verifiable Identities

Humans and organizations interacting over digital networks already rely extensively on identity,

authentication and authorization. AI agents should not become an exception.

The discussion document itself raises questions concerning agent identity, detailed activity records, human oversight and chain of custody when agents interact with one another. (Discussion Document, p. 29.) This becomes especially important when an agent can spend money, obtain confidential data or manipulate real-world systems.

Canada should support interoperable agent credentials capable of communicating, where appropriate, the responsible organization or user; the agent or application involved; the authority delegated to it; its relevant certification or attestation status; and a transaction or session identifier.

This does not mean that every harmless AI process requires government registration. Requirements should be proportional to capability and risk. But an AI agent authorized to transact with a bank, interact with government infrastructure or control an industrial system should be distinguishable from an unidentified process arriving over the public Internet.

Services operating in higher-risk environments should also be able to require verifiable agent credentials and reject unidentified automated actors. This could eventually provide a foundation not only for transparency but for machine-to-machine trust.

6. Incident Reporting Requires Evidence

The discussion document makes a particularly useful comparison between AI incident reporting and safety-reporting systems in sectors such as aviation. It recognizes a difficult incentive problem: if every admission of an incident immediately creates liability, organizations may become reluctant to report problems. It therefore considers combinations of mandatory reporting, confidential reporting and no-fault mechanisms that allow safety lessons to be shared. (Discussion Document, pp. 22–24.)

I strongly support developing such mechanisms. However, an incident-reporting regime will only be as useful as the evidence available to investigators. This is why incident reporting and auditability should be developed together.

When a serious AI incident occurs, investigators should ideally be able to establish: What system was operating? Which model and software versions were involved? What permissions did the system possess? Who authorized those permissions? What tools, data and external systems did it access? What actions did it take? What occurred immediately before the failure? When did humans intervene? Could the incident reasonably have been prevented or reproduced?

Without adequate technical records, incident reporting risks becoming little more than competing narratives after something has already gone wrong. The objective should be to create enough trustworthy evidence to learn from failures.

7. Regulate Where Canada Has Practical Leverage

Artificial intelligence is inherently international. Canadians can access models developed, trained and hosted almost anywhere in the world. Open models can also be downloaded, modified and redeployed by organizations over which Canadian authorities may have little practical control.

Canada should therefore be cautious about regulatory approaches that assume it can directly govern the internal design or training process of every AI system Canadians might encounter. The discussion document itself recognizes the complexity of the AI value chain: developers, deployers and users may be separate

parties with only partial visibility into one another's activities. (Discussion Document, p. 7.)

Canada has considerably greater leverage over deployment. That includes AI systems procured by Canadian governments, systems deployed commercially by Canadian organizations, regulated industries, critical infrastructure, services offered to Canadians and the interfaces through which AI agents interact with Canadian systems.

This suggests emphasizing practical mechanisms such as certification for high-impact applications; standardized testing; human recourse; auditability; agent identification; procurement requirements; incident reporting; and common technical standards.

Canada's broader AI strategy already emphasizes sovereign infrastructure, domestic AI capacity and international standards cooperation. Transparency policy can complement those efforts rather than attempting to substitute for them.

Human Administrator Rights

These recommendations can be summarized through a simple systems metaphor. As software engineers, we do not normally build a critical system and then intentionally remove every administrator account. We preserve the ability to inspect the system, determine what happened, change permissions, stop unsafe processes and recover when something fails.

As increasingly capable artificial intelligence becomes embedded in institutions and infrastructure, society should retain analogous controls. I suggest organizing these Human Administrator Rights as a three-layer Rights Stack.

01 HUMAN AUTHORITY

Humans remain sovereign over consequential machines.

KNOW · READ · APPEAL · OVERRIDE · AUDIT · STOP · OFFLINE

These are the operational protections at the heart of this submission: know when AI materially affects you; understand consequential systems; challenge important automated decisions; preserve accountable human override; maintain trustworthy audit records; retain the ability to stop unsafe systems; and preserve workable non-AI fallbacks for essential functions.

02 HUMAN PARTICIPATION

Automation must benefit the people and societies that make it possible.

SHARE · OWN · FAIR APPLICATION

These principles address the distribution of automation's benefits and burdens: people should share materially in productivity gains; intelligence indispensable to public life should not become permanently and unaccountably monopolized; and the costs of automation should not be imposed on some groups while others are structurally exempted from the same transition.

03 HUMAN INTEGRITY

Technology must not take ownership of the person.

PRIVACY · IDENTITY · CONSENT · REMAIN HUMAN

These principles address personal autonomy: protection of personal, behavioural and neural data; meaningful control over likeness, voice and digital identity; informed consent for artificial, genetic or neural

augmentation; and the right not to be compelled to augment one's body or mind merely to participate fully in society.

The first layer is directly relevant to this transparency consultation. The second and third layers are broader principles for future policy consideration as AI increasingly affects economic participation, identity, privacy and human autonomy.

AI does not replace human rights. It requires us to extend them.

Conclusion

Canada's discussion document asks the right initial question: what transparency do Canadians need as artificial intelligence becomes more deeply embedded in everyday life?

My suggestion is to carry that question one step further. Transparency should not merely allow someone to discover that an AI system exists. It should allow people, organizations, investigators and regulators to determine what acted; under whose authority; using which permissions and information; according to which rules; what actions resulted; and what can be done when something goes wrong.

That is the distinction between disclosure and control.

In the proposed Rights Stack, these are primarily Human Authority protections. Human Participation and Human Integrity extend the same principle into the economic and personal domains as AI capabilities continue to expand.

Canada has an opportunity to help establish practical standards for independent testing, provenance interfaces, human recourse, agent identity, auditability and incident reconstruction without attempting to micromanage the internal design of every artificial-intelligence system in the world.

In my view, that would be a useful Canadian contribution to international AI governance.

Artificial intelligence can be given enormous capabilities and significant autonomy. Humans should retain administrator access.

About the Submitter

I am a Canadian software architect and systems engineer with more than 25 years of experience designing and building production software, databases, infrastructure and automated systems.

I am submitting these comments in my personal capacity.

I am not an AI-policy scholar, lawyer or representative of an organization. My perspective is primarily that of a working technologist accustomed to thinking about systems in terms of reliability, permissions, logging, failure recovery, accountability and human control.

Artificial-intelligence tools assisted me in researching, organizing, challenging and editing this submission. I reviewed the arguments, selected the recommendations and accept responsibility for the final submission.

Reference

Innovation, Science and Economic Development Canada. Enhancing Trust in Artificial Intelligence Through Increased Transparency. Discussion Document, July 2026.