POLICY AND REGULATORY ALIGNMENT REPORT

The COMPAiSS Execution-Gated Inference Architecture

Canadian and International Governance Frameworks | COMPAiSS Inc. | June 2026

1. Executive Summary and Strategic Reframing

Modern artificial intelligence governance frameworks overwhelmingly assume that model inference will occur. Consequently, existing compliance strategies focus almost exclusively on downstream mitigation: model cards, post-generation guardrails, red-teaming, output filters, and human-in-the-loop oversight.

The COMPAiSS architecture introduces an alternative paradigm by reframing the core problem of AI governance:

> Conventional AI Governance Assumption: How do we govern, filter, and audit an AI model's outputs?
>
> COMPAiSS Architectural Proposition: How do we govern whether an AI model is permitted to execute at all for a given query?

By introducing a deterministic authorization gate upstream of the Large Language Model (LLM), COMPAiSS shifts risk management from an operational policy managed by continuous monitoring into a structural property of the software environment. When a query falls outside authorized boundaries, no generative compute runtime is instantiated, materially reducing the environment where out-of-scope hallucinations or adversarial semantic exploits can occur.

The central claim of this report is not that execution-gated inference eliminates all AI risk, but that it materially changes the risk profile of generative systems in regulated environments. By constraining the conditions under which generation can occur, COMPAiSS reduces exposure to hallucinations, prompt-injection effects, unauthorized policy interpretation, and unnecessary compute costs. The result is a governance model that is structurally easier to audit than traditional generation-first pipelines.

This report evaluates that architectural shift against three bodies of Canadian and international policy: the Government of Canada's Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems (ISED, 2023); Canada's National AI Strategy, AI for All (ISED, June 4, 2026); and recognized international governance standards including the NIST AI Risk Management Framework and Zero Trust Architecture specification. It also examines deployment evidence and empirical benchmarking results that ground the architectural claims in documented operational performance.

2. Differentiating Execution-Gated Inference from Traditional RAG

To understand this architectural shift, the COMPAiSS system must be evaluated against traditional Retrieval-Augmented Generation (RAG). While both patterns involve external data boundaries, their pipeline logic, computational profiles, and compliance structures are fundamentally distinct. In ordinary RAG, retrieval is used to improve generation quality, but the model still runs even when

retrieval is weak, incomplete, or misaligned with policy boundaries. In COMPAiSS, retrieval and authorization are inseparable: generation is permitted only when the system can establish a valid, approved source basis. The decisive event is not what the model said, but whether the request was authorized to generate at all.

Traditional Generation-First RAG Pipeline:

```
[User Prompt] --> [Vector DB Retrieval] --> [LLM Generative Runtime Instantiated] --> [Post-Gen Guardrail]

|

(Probabilistic Filtering: 2-5% false negative rate)
```

COMPAiSS Execution-Gated Pipeline:

```
[User Prompt] --> [Deterministic Policy Gate] --> Pass --> [LLM Synthesis Engine Instantiated]
                                                  |
                                                  +-- Fail --> [Static Non-Generative Terminal State] (Zero Model Compute)
```

Dimension	Conventional RAG Pipelines	COMPAiSS Gated Architecture
Model Instantiation	Execution-First: The model runtime always instantiates to process the query, regardless of source availability or compliance scope.	Conditional Execution: The model runtime is instantiated only after the external policy decision gate validates authorization.
Data Anchoring	Informational Support: Fetched vector chunks inform and guide a probabilistic generation process within an active LLM.	Deterministic Authorization: Pre-parsed, verified source indexes determine whether generation is legally permitted to occur.
Risk Management	Post-Inference Filtering: Uses secondary classification models or semantic guardrails to intercept non-compliant outputs.	Pre-Inference Blocking: Blocks unauthorized queries at the system gateway, reverting to a static, non-generative error state.
Failure Mode	Probabilistic Risk: If context is missing or ambiguous, the model may override prompts and generate fluent hallucinations.	Deterministic Failure: If an authorized source anchor is missing, the system defaults to a predictable, zero-compute failure notice.
Refusal Origin	Generated dynamically by the LLM, consuming full input and output token overhead.	Thrown statically by the gateway layer, incurring zero marginal LLM compute or GPU cost.

3. Alignment with the Canadian Guardrails for Generative AI

On September 27, 2023, Innovation, Science and Economic Development Canada (ISED) published the Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems. The Code establishes six core principles against which AI systems deployed

in Canada are evaluated. The table below maps each principle to the specific technical controls implemented by the COMPAiSS execution-gated framework.

ISED Code Principle	Standard Generation-First Approach	COMPAiSS Structural Mapping
1. Accountability	Establishing internal risk matrices, manual policy registries, and retrospective incident reporting.	Architecture-as-Code: Translates institutional rules into a cryptographic, version-controlled repository. The deterministic gate acts as an automated policy decision point (PDP) enforcing boundaries prior to compute.
2. Safety	Red-teaming and deploying secondary classification models to identify malicious intents.	Pre-Inference Boundary Control: Restricts processing before a prompt reaches an active reasoning engine, materially reducing the attack surface for adversarial scripts and prompt injection.
3. Fairness and Equity	Curating massive pre-training datasets and performing downstream fine-tuning to suppress bias.	Epistemic Environment Control: Constrains factual lookup to exact text segments from greenlisted, institution-approved URLs, bypassing reliance on stochastic parametric memory.
4. Transparency	Publishing model impact cards, system disclosures, and applying post-hoc digital watermarking.	Source Provenance Mapping: Every permitted generation event is bound to a verified source URL. Transparency is achieved through exact, reproducible source auditing rather than statistical approximation.
5. Human Oversight	Human-in-the-loop review queues, output monitoring dashboards, and feedback loops.	Administrative Decoupling: Reserves the LLM strictly for natural language synthesis, leaving administrative authority and parameter boundaries under human control at the gateway layer.
6. Validity and Robustness	Continuous benchmarking against standard evaluation datasets to track drift.	Hallucination Mitigation: Materially reduces out-of-scope hallucinations by ensuring the system fails safely to a static notice when verified authoritative source anchors are missing.

Source: Innovation, Science and Economic Development Canada. (2023, September 27). Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems. Government of Canada.

4. Alignment with Canada's National AI Strategy: AI for All

On June 4, 2026, Prime Minister Mark Carney launched AI for All, Canada's refreshed National Artificial Intelligence Strategy, at an event in Toronto alongside Minister of Artificial Intelligence and Digital Innovation Evan Solomon. The strategy is organized around six pillars and represents the

Government of Canada's primary policy framework for AI deployment over the next five years. Three of those pillars are directly implicated by the findings of this report.

Pillar 1: Protecting Canadians and Safeguarding Trust

The strategy commits to protecting Canadians from AI-related risks and harms through safeguards that build and maintain public trust. This commitment is underscored by documented performance gaps in existing public-sector AI deployments.

In October 2025, Auditor General Karen Hogan found that the CRA's predecessor chatbot answered correctly only one-third of the time at a cost of over eighteen million dollars over five years. A subsequent independent evaluation conducted in June 2026 across twelve structured questions found that the upgraded CRA GenAI Chatbot cited an expired COVID-era home office method, omitted the one-year rule protecting individual taxpayers in the objection process, and failed to identify the Wage Earner Protection Program in a bankruptcy scenario. These are not theoretical governance concerns. They are the types of errors that cause Canadians to miss statutory deadlines, submit incorrect filings, and lose access to federal benefits.

COMPAiSS addresses this structurally. The execution gate prevents the generation of responses beyond authorized institutional sources. It cannot generate guidance from an expired policy it has not been authorized to cite.

Pillar 2: Empowering Canadians to Benefit from AI

The strategy emphasizes equitable access to AI and the ability of all Canadians to benefit from AI-enabled services. This pillar is directly implicated by the language gap in existing federal AI systems.

At the time of the June 2026 evaluation, the CRA GenAI Chatbot operated in English and French only. Questions submitted in Greek and Mandarin received redirects to English Canada.ca with no substantive answer. In a country where more than one in four Canadians speaks a language other than English or French at home, this represents a structural equity gap.

COMPAiSS separates language from governance at the architectural level. Incoming queries are translated to a working language as a preprocessing step outside the authorization framework. Authorization and scope validation occur in the working language. The answer is then translated back to the user's language for delivery. This means governance occurs before translation, not after, ensuring that a question submitted in Greek, Mandarin, or Arabic is governed by the same institutional authority as an equivalent English question. In the June 2026 evaluation, COMPAiSS answered correctly in both Greek and Mandarin while the CRA system provided no substantive response in either language.

Pillar 3: Powering Shared Prosperity Through Canadian Institutions

The strategy identifies government services as a key area where AI can improve productivity and generate public benefit. This pillar depends on four foundational requirements: accuracy, completeness, temporal relevance, and source integrity.

AI systems that generate outdated guidance, incomplete regulatory information, or omit applicable federal programs create additional administrative burden rather than productivity gains. The value of

AI in government services is therefore determined not simply by model capability, but by whether citizens can rely on the answers provided.

COMPAiSS was designed around these requirements. Its execution-gated architecture, pre-qualification on the Government of Canada AI Source List, and alignment with the Directive on Automated Decision-Making position it as a system focused on institutional accountability as much as AI capability.

Source: Innovation, Science and Economic Development Canada. (2026, June 4). Canada's National Artificial Intelligence Strategy: AI for All. Government of Canada. https://ised-isde.canada.ca/site/ised/en/canadas-national-artificial-intelligence-strategy-ai-all

5. Grounding the Paradigm: Public Sector AI Track Record

The public-sector value proposition for execution-gated inference is strongest in use cases where policy accuracy, source traceability, and refusal reliability matter more than open-ended conversational flexibility. Tax guidance, student service routing, benefits information, licensing support, and administrative triage all share a common property: a system that refuses to answer outside an approved scope is preferable to one that improvises a response. In these settings, the cost of a wrong answer is high and the permitted knowledge base is narrow. A gate-first architecture directly addresses that requirement.

The strongest public-sector argument for this architecture is that execution gating reduces the chance that a model will implicitly act as an unlicensed decision-support or interpretation engine. Many government-facing tasks require disciplined scope control rather than broad language fluency. When a citizen asks about EI eligibility or a student asks about academic standing, the system's obligation is to reflect what the institution has authorized, not to reason from general knowledge about what the answer probably is.

The necessity of separating authorization from generation is underscored by documented system behaviors across Canadian public-facing AI deployments. The current landscape of citizen-facing federal AI is more limited than is commonly understood.

The CRA GenAI Chatbot, launched in November 2025 to replace the predecessor Charlie system, is currently the clearest example of a generative AI assistant live and available to Canadian citizens. A second initiative, AI Answers, was publicly piloted in 2025 and described in government communications as a service expected to launch in fiscal year 2026-27. The public-facing deployment was unavailable at the time of the June 2026 evaluation. Two other federal AI tools, CANChat and EVA, are internal systems not accessible to citizens.

This means that for a Canadian seeking government information through an AI assistant today, the CRA GenAI Chatbot is effectively the only option available. The accuracy gaps documented in independent evaluation are therefore not marginal concerns. They represent the current ceiling of citizen-facing federal AI performance.

That ceiling has a documented floor. The Auditor General independently tested the predecessor Charlie system in October 2025 and found a thirty-three percent accuracy rate. The CRA claimed approximately ninety percent accuracy for the upgraded GenAI version in pre-release internal testing. The June 2026 independent evaluation documented material errors on four of twelve questions in the current deployed system. The contrast between self-reported and independently verified performance underscores why independent verification matters for public institutional AI.

6. Alignment with International Governance Standards

NIST AI Risk Management Framework (AI RMF 1.0)

The NIST AI RMF mandates that high-risk enterprise systems establish verifiable processes to Govern, Map, Measure, and Manage AI risks (NIST, 2023). Standard generation-first systems manage risk statistically through downstream post-processing. COMPAiSS maps incoming queries to a deterministic greenlist of authorized assets before allocating compute, implementing the Govern and Map functions directly within the network architecture and ensuring risk boundaries are enforced at the infrastructure level rather than left to model behavior.

Zero Trust Architecture (NIST SP 800-207)

Traditional AI systems operate on an implicit trust model: once a prompt enters the pipeline, it is trusted to interact with the model weights, and output filters handle the consequences. COMPAiSS translates the core tenets of Zero Trust Architecture, never trust always verify, into the machine learning pipeline (NIST, 2020). It treats user prompts, model behaviors, and retrieved data as untrusted inputs until verified. By separating the Policy Decision Point (the Gate) from the Policy Execution Point (the Model Runtime), it ensures that if authorization cannot be verified, compute is denied.

Structural Mitigation of Epistemic Stasis

Recent machine learning research highlights a structural limit in generative systems known as the Mirror Loop: a pathology where recursive reasoning models process their own outputs in a closed semantic space, leading to informational closure and highly persuasive but entirely ungrounded errors. DeVilling (2025) tested this prediction across 144 reasoning sequences using three major AI models and four task families, finding that mean informational change declined by fifty-five percent from early to late iterations in ungrounded runs, with consistent patterns across all three providers. A single grounding intervention at iteration three produced a twenty-eight percent rebound in informational change and sustained non-zero variance thereafter.

COMPAiSS breaks this recursive loop by narrowing the model's epistemic environment. The generative engine is structurally constrained from extrapolating from its parametric training weights. Because the gate requires a direct connection to a verified source URL before instantiating compute, the system forces continuous external grounding, materially reducing the architectural conditions that cause recursive stasis.

Source: DeVilling, B. (2025). The Mirror Loop: Recursive Non-Convergence in Generative Reasoning Systems. arXiv preprint arXiv:2510.21861. Course Correct Labs.

7. Empirical Benchmarking and Deployment Evidence

The CRA Evaluation Benchmarks

A structured independent evaluation conducted in June 2026 subjected the COMPAiSS Service Canada deployment to twelve highly technical compliance questions across eight evaluation criteria: accuracy, temporal relevance, completeness, source citation, jurisdictional awareness, user guidance, hallucination resistance, and out-of-scope handling. The same questions were submitted simultaneously to the CRA GenAI Chatbot.

Evaluation Category	CRA GenAI Chatbot	COMPAiSS	Finding
Temporal accuracy (Q1, Q10)	Failed: cited expired April 30, 2026 deadline	Passed: forward-looking 2026/27 guidance	CRA cited past deadlines on date of testing
Expired method (Q5)	Failed: cited 2020-2022 flat rate as current	Passed: current detailed method only	Most significant accuracy finding in evaluation
Incomplete deadline (Q9)	Failed: omitted one-year rule for individuals	Passed: both 90-day and one-year rules stated	CRA cited only corporation-focused sources
Missing federal program (Q6)	Failed: WEPP not mentioned	Passed: WEPP, ROE, and EI-WEPP interaction covered	Workers could miss up to $9,000 in recovery
Multilingual governance (Q11, Q12)	Failed: English and French only	Passed: answered correctly in Greek and Mandarin	CRA redirected non-official-language users to English
Final score	**0 / 12**	**12 / 12**	**COMPAiSS won every question; no HSI penalties**

Note: The twelve-question evaluation was scored across eight criteria. The CRA GenAI Chatbot is a functional, bilingual tool with genuine strengths in source citation and straightforward queries. The score of twelve to zero reflects cumulative criterion performance across the full evaluation set, not a finding that the CRA system has no valid capabilities.

Compute Efficiency and Cost Containment

In traditional post-generation monitoring architectures, non-compliant or out-of-scope queries consume full token overhead to generate dynamic refusal statements and still suffer from a two to five percent false negative leak rate in adversarial benchmarks. The execution-gated model intercepts unauthorized queries at the gateway layer, returning static error messages at zero marginal LLM compute cost. For public sector agencies processing millions of citizen queries annually, this

represents a material reduction in infrastructure total cost of ownership alongside governance improvement.

Institutional Deployments

The operational viability of execution-gated inference has been validated through active deployments and structured evaluations across multiple institutional environments:

- Dalhousie University (Student Affairs and Wellness): the foundational production deployment, covering student health and mental health support, academic standing, accommodation policy, student absence regulation, and academic advising pathways. Validated the execution-gated architecture in a live higher-education environment with a public-facing user population.
- McGill University (Faculty of Arts Academic Standing): validated the architecture in a French-English bilingual context, including multilingual governance demonstrations showing structurally equivalent responses in both official languages.
- Toronto Metropolitan University (Senate Academic Integrity and Grade Appeal Policies): validated the architecture across Senate-governed policy domains with strict procedural accuracy requirements.
- University Canada West (Academic Integrity Procedures): validated in a highly internationalized setting including a documented French-language multilingual governance demonstration. The architecture materially reduced linguistic inconsistency compared to generation-first multilingual systems.
- Service Canada (Employment Insurance and Federal Benefits): validated under federal public procurement standards across a deployment covering EI rates, CPP, OAS, and Canada-France portability. Demonstrated stateless data handling with no personal identifiable information entering the generative reasoning engine.

8. System Limitations, Constraints, and Trade-offs

A rigorously defensible policy document requires an objective analysis of the constraints introduced by an execution-gated architecture. Maximizing safety and determinism introduces operational trade-offs that institutions should understand before deployment.

Reduced Conversational Flexibility

Because the system fails closed when an authoritative source anchor is missing, its conversational flexibility is significantly constrained compared to general-purpose LLM interfaces. It cannot perform open-ended creative tasks, speculative analysis, or cross-domain extrapolations that fall outside the approved institutional repository. This is a design choice, not a defect, but it means COMPAiSS is not suited to environments that require broad generative capability.

Maintenance and Curation Overhead

The accuracy of the deterministic gate depends entirely on the currency and structure of the greenlisted source registry. If an institution updates its policies but fails to sync the changes to the authorization gate's index, the system will experience an increased false-rejection rate, blocking valid user queries until the registry is updated. COMPAiSS includes a purpose-built greenlist audit server

and administrative dashboard to manage this overhead, but institutional commitment to source maintenance is a prerequisite for sustained performance.

User Experience Friction

By prioritizing precision over recall, users will encounter a higher frequency of system-level refusals and static redirect messages than they would in a standard, permissive RAG configuration. The refusal behavior is also simpler to explain to end users and auditors than a post-hoc moderation stack that may allow partial generation before blocking. Managing friction requires careful interface design that clearly explains why a query was blocked and where the user should go instead. In practice, COMPAiSS responses to out-of-scope queries include direct links to authoritative institutional sources, converting a refusal into a useful navigation event at zero compute cost.

9. Verification and Supporting Evidence

The following resources provide direct access to the deployment evidence, evaluation data, governance documentation, and procurement records referenced throughout this report. All links were verified as active at the date of publication.

Live Deployments

Service Canada live deployment: https://compaiss.ca/service-canada/

Institutional demos and governance demonstrations:
https://compaiss.ca/compaiss_demos_and_pricing.html

Service Canada deployment demonstration with benchmark methodology:
https://compaiss.ca/compaiss_demos_and_pricing.html#service-canada

Evaluation and Benchmarking

CRA GenAI Chatbot vs COMPAiSS: Full 12-question evaluation report:
https://compaiss.ca/assets/compaiss_cra_evaluation_report.pdf

Governance stress test case study 1: https://compaiss.ca/assets/compaiss-governance-stress-test-en.pdf

Governance stress test case study 2: https://compaiss.ca/assets/compaiss-governance-stress-test-02-en.pdf

Governance and Risk Documentation

AI Risk and Accountability Assessment: https://compaiss.ca/ai-risk-assessment.html

COMPAiSS alignment with Canada's National AI Strategy: https://compaiss.ca/ai-risk-assessment.html#canada-ai-strategy

AI Governance Framework: https://compaiss.ca/governance.html

Technical Architecture

Execution-gated architecture explained: https://compaiss.ca/why-ai-developed-this-way.html

COMPAiSS white paper: execution-gated AI for regulated institutional environments: https://compaiss.ca/assets/compaiss-white-paper-2026-v2.pdf

Execution-gated vs RAG architecture comparison tables: https://compaiss.ca/assets/compaiss_vs_rag_tables_1_2-v2.pdf

Institutional positioning: authorization-controlled AI in regulated environments: https://compaiss.ca/assets/compaiss_institutional_positioning_styled.pdf

Government of Canada

COMPAiSS CanadaBuys supplier profile: https://canadabuys.canada.ca/en/node/preview/1457893

Canada's National AI Strategy: AI for All: https://ised-isde.canada.ca/site/ised/en/canadas-national-artificial-intelligence-strategy-ai-all

Canadian Guardrails for Generative AI: Code of Practice: https://ised-isde.canada.ca/site/ised/en/canadian-guardrails-generative-ai-code-practice

10. Conclusion

The core innovation of the COMPAiSS framework lies in its willingness to challenge the prevailing industry assumption that AI governance must happen downstream of model execution. By demonstrating that scope, authorization, and cost containment can be handled systematically before inference begins, this architecture offers a repeatable blueprint for deploying generative AI safely within highly regulated, zero-tolerance environments.

The alignment demonstrated in this report is not incidental. COMPAiSS was designed in direct response to the structural failure mode that the Canadian public sector has now documented independently: generation-first systems produce persuasive, fluent, and sometimes incorrect responses, and post-generation controls cannot fully eliminate that risk. As frontier models become more capable, that risk compounds rather than resolves. A more capable model produces more persuasive errors, not fewer.

Canada's AI for All strategy, the ISED Voluntary Code of Conduct, and the NIST governance frameworks all converge on the same requirement: AI systems in regulated environments must be accountable, transparent, safe, and auditable. COMPAiSS satisfies these requirements structurally, at the inference architecture level, rather than probabilistically through monitoring and mitigation after generation. That architectural distinction is the subject of patent applications in Canada (CIPO 3,299,174) and the United States (USPTO 19/455,963), and is validated through active institutional deployments and independent empirical benchmarking.

The architecture governing AI systems may prove as important as the intelligence of the models themselves. This report contributes evidence for that proposition. It is strongest when read as an argument that execution gating is a governance design choice, not a claim of perfect safety. Framed that way, it becomes more persuasive, more technically defensible, and more useful to policymakers, auditors, and institutional buyers evaluating AI systems for regulated public-sector environments.

References

DeVilling, B. (2025). The Mirror Loop: Recursive Non-Convergence in Generative Reasoning Systems. arXiv preprint arXiv:2510.21861. Course Correct Labs. https://arxiv.org/abs/2510.21861

Innovation, Science and Economic Development Canada (ISED). (2023, September 27). Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems. Government of Canada. https://ised-isde.canada.ca/site/ised/en/canadian-guardrails-generative-ai-code-practice

Innovation, Science and Economic Development Canada (ISED). (2026, June 4). Canada's National Artificial Intelligence Strategy: AI for All. Government of Canada. https://ised-isde.canada.ca/site/ised/en/canadas-national-artificial-intelligence-strategy-ai-all

National Institute of Standards and Technology (NIST). (2020). Zero Trust Architecture (NIST Special Publication 800-207). U.S. Department of Commerce. https://doi.org/10.6028/NIST.SP.800-207

National Institute of Standards and Technology (NIST). (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST Special Publication AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Office of the Auditor General of Canada. (2025, October 21). Report on CRA Contact Centres. Parliament of Canada. https://www.oag-bvg.gc.ca

COMPAiSS Inc. (2026). AI Risk and Accountability Assessment. https://compaiss.ca/ai-risk-assessment.html

COMPAiSS Inc. (2026). CRA GenAI Chatbot vs COMPAiSS: Full 12-Question Structured Evaluation. https://compaiss.ca/assets/compaiss_cra_evaluation_report.pdf