RAPPORT D'ALIGNEMENT RÉGLEMENTAIRE ET POLITIQUE

L'architecture à inférence conditionnelle de COMPAiSS

Cadres canadiens et internationaux de gouvernance | COMPAiSS Inc. | Juin 2026

1. Résumé exécutif et recentrage stratégique

Les cadres contemporains de gouvernance de l'intelligence artificielle supposent en grande majorité que l'inférence du modèle aura lieu. En conséquence, les stratégies de conformité existantes se concentrent presque exclusivement sur l'atténuation en aval : fiches descriptives des modèles, garde-fous post-génération, séances de mise à l'épreuve, filtres de sortie et supervision humaine en boucle.

L'architecture COMPAiSS introduit un paradigme alternatif en reformulant le problème fondamental de la gouvernance de l'IA :

> Hypothèse conventionnelle de gouvernance de l'IA : Comment gouverner, filtrer et auditer les résultats produits par un modèle d'IA ?
>
> Proposition architecturale de COMPAiSS : Comment déterminer si un modèle d'IA est autorisé à s'exécuter pour une requête donnée ?

En introduisant un portail d'autorisation déterministe en amont du grand modèle de langage (GML), COMPAiSS fait passer la gestion des risques d'une politique opérationnelle assurée par une surveillance continue à une propriété structurelle de l'environnement logiciel. Lorsqu'une requête dépasse les limites autorisées, aucun environnement d'exécution génératif n'est instancié, réduisant matériellement le contexte dans lequel des hallucinations hors portée ou des exploits sémantiques adversariaux peuvent survenir.

L'argument central de ce rapport n'est pas que l'inférence à exécution conditionnelle élimine tous les risques liés à l'IA, mais qu'elle modifie matériellement le profil de risque des systèmes génératifs dans les environnements réglementés. En restreignant les conditions dans lesquelles la génération peut avoir lieu, COMPAiSS réduit l'exposition aux hallucinations, aux effets d'injection de commandes, à l'interprétation non autorisée de politiques et aux coûts de calcul superflus. Il en résulte un modèle de gouvernance structurellement plus facile à auditer que les chaînes de traitement classiques fondées sur la génération en premier.

Le présent rapport évalue ce changement architectural au regard de trois corpus de politiques canadiennes et internationales : le Code de conduite volontaire du gouvernement du Canada sur le développement et la gestion responsables des systèmes d'IA générative avancés (ISDE, 2023) ; la Stratégie nationale en matière d'IA du Canada, L'IA pour tous (ISDE, 4 juin 2026) ; et les normes internationales reconnues de gouvernance, dont le Cadre de gestion des risques liés à l'IA du NIST et la spécification d'architecture à vérification systématique. Il examine également les données probantes sur les déploiements et les

résultats d'évaluation comparative empirique qui ancrent les affirmations architecturales dans une performance opérationnelle documentée.

2. Distinction entre l'inférence à exécution conditionnelle et la RAG traditionnelle

Pour bien saisir ce changement architectural, le système COMPAiSS doit être évalué en comparaison avec la génération augmentée par récupération (RAG) traditionnelle. Bien que les deux modèles impliquent des frontières de données externes, leur logique de chaîne de traitement, leurs profils de calcul et leurs structures de conformité sont fondamentalement distincts. Dans la RAG ordinaire, la récupération sert à améliorer la qualité de la génération, mais le modèle s'exécute quand même lorsque la récupération est faible, incomplète ou mal alignée avec les limites de la politique. Dans COMPAiSS, récupération et autorisation sont indissociables : la génération n'est permise que lorsque le système peut établir une base de sources validées et approuvées. L'événement décisif n'est pas ce que le modèle a produit, mais si la requête était autorisée à générer quoi que ce soit.

Chaîne de traitement RAG classique (génération en premier) :

```
[Requête utilisateur] --> [Récupération base vectorielle] --> [Exécution générative GML instanciée] --> [Garde-fou post-génération]

|

(Filtrage probabiliste : taux de faux négatifs de 2 à 5 %)
```

Chaîne de traitement COMPAiSS à exécution conditionnelle :

```
[Requête utilisateur] --> [Portail de politique déterministe] --> Autorisé --> [Moteur de synthèse GML instancié]
                                |
                                +-- Refus --> [État terminal statique non génératif] (Zéro calcul de modèle)
```

Dimension	Chaînes RAG conventionnelles	Architecture à portail COMPAiSS
Instanciation du modèle	Exécution en premier : l'environnement d'exécution du modèle s'instancie toujours pour traiter la requête, indépendamment de la disponibilité des sources ou du périmètre de conformité.	Exécution conditionnelle : l'environnement d'exécution du modèle n'est instancié qu'après que le portail externe de décision de politique a validé l'autorisation.
Ancrage des données	Appui informationnel : les fragments vectoriels récupérés informent et orientent un processus de génération probabiliste au sein d'un GML actif.	Autorisation déterministe : les index de sources préanalysés et vérifiés déterminent si la génération est légitimement autorisée.

Gestion des risques	Filtrage post-inférence : utilise des modèles de classification secondaires ou des garde-fous sémantiques pour intercepter les résultats non conformes.	Blocage pré-inférence : bloque les requêtes non autorisées au portail du système et revient à un état d'erreur statique non génératif.
Mode de défaillance	Risque probabiliste : si le contexte est absent ou ambigu, le modèle peut outrepasser les instructions et produire des hallucinations fluentes.	Défaillance déterministe : si une ancre de source autorisée est absente, le système bascule vers un avis de défaillance prévisible à zéro calcul.
Origine du refus	Généré dynamiquement par le GML, consommant la totalité des jetons d'entrée et de sortie.	Émis statiquement par la couche du portail, sans coût marginal de calcul GML ou GPU.

3. Alignement avec les garde-fous canadiens pour l'IA générative

Le 27 septembre 2023, Innovation, Sciences et Développement économique Canada (ISDE) a publié le Code de conduite volontaire sur le développement et la gestion responsables des systèmes d'IA générative avancés. Le Code établit six principes fondamentaux à l'aune desquels sont évalués les systèmes d'IA déployés au Canada. Le tableau ci-dessous fait correspondre chaque principe aux contrôles techniques spécifiques mis en œuvre par le cadre à exécution conditionnelle de COMPAiSS.

Principe du Code ISDE	Approche classique de génération en premier	Correspondance structurelle COMPAiSS
1. Responsabilité	Établissement de matrices de risques internes, de registres de politiques manuels et de rapports d'incidents rétrospectifs.	Architecture-en-tant-que-code : traduit les règles institutionnelles en un répertoire cryptographique sous contrôle de version. Le portail déterministe joue le rôle de point de décision de politique automatisé (PDP) appliquant les limites avant tout calcul.
2. Sécurité	Séances de mise à l'épreuve et déploiement de modèles de classification secondaires pour identifier les intentions malveillantes.	Contrôle des limites pré-inférence : restreint le traitement avant qu'une instruction n'atteigne un moteur de raisonnement actif, réduisant matériellement la surface d'attaque pour les scripts adversariaux et l'injection d'instructions.
3. Équité et justice	Conservation de vastes ensembles de données de préentraînement et ajustement fin en aval pour atténuer les biais.	Contrôle de l'environnement épistémique : limite la recherche factuelle aux segments de texte exacts provenant d'URL institutionnelles approuvées figurant sur la liste verte, contournant le recours à la mémoire paramétrique stochastique.

4. Transparence	Publication de fiches d'impact des modèles, divulgations système et application de filigranes numériques post-hoc.	Cartographie de la provenance des sources : chaque événement de génération autorisé est lié à une URL de source vérifiée. La transparence est atteinte par un audit de source exact et reproductible plutôt que par approximation statistique.
5. Surveillance humaine	Files de révision avec intervention humaine, tableaux de bord de surveillance des sorties et boucles de rétroaction.	Découplage administratif : réserve le GML strictement à la synthèse en langage naturel, laissant l'autorité administrative et les limites paramétriques sous contrôle humain au niveau de la couche du portail.
6. Validité et robustesse	Évaluation comparative continue par rapport aux ensembles de données d'évaluation normalisés pour suivre la dérive.	Atténuation des hallucinations : réduit matériellement les hallucinations hors portée en garantissant que le système bascule vers un avis statique lorsque les ancres de sources faisant autorité vérifiées sont absentes.

Source : Innovation, Sciences et Développement économique Canada. (2023, 27 septembre). Code de conduite volontaire sur le développement et la gestion responsables des systèmes d'IA générative avancés. Gouvernement du Canada.

4. Alignement avec la Stratégie nationale en matière d'IA du Canada : L'IA pour tous

Le 4 juin 2026, le premier ministre Mark Carney a lancé L'IA pour tous, la Stratégie nationale renouvelée en matière d'intelligence artificielle du Canada, lors d'un événement à Toronto aux côtés du ministre de l'Intelligence artificielle et de l'Innovation numérique, Evan Solomon. La stratégie est organisée autour de six piliers et représente le principal cadre de politique du gouvernement du Canada pour le déploiement de l'IA au cours des cinq prochaines années. Trois de ces piliers sont directement impliqués par les conclusions du présent rapport.

Pilier 1 : Protéger les Canadiens et préserver la confiance

La stratégie s'engage à protéger les Canadiens contre les risques et préjudices liés à l'IA grâce à des garanties qui bâtissent et maintiennent la confiance du public. Cet engagement est souligné par les lacunes de performance documentées dans les déploiements d'IA existants dans le secteur public.

En octobre 2025, la vérificatrice générale Karen Hogan a constaté que le clavardeur prédécesseur de l'ARC répondait correctement seulement un tiers du temps, à un coût de

plus de dix-huit millions de dollars sur cinq ans. Une évaluation indépendante ultérieure menée en juin 2026 sur douze questions structurées a révélé que le Clavardeur GenAI mis à niveau de l'ARC citait une méthode de bureau à domicile de l'ère COVID expirée, omettait la règle d'un an protégeant les contribuables particuliers dans le processus d'opposition, et ne parvenait pas à identifier le Programme de protection des salariés dans un scénario de faillite. Il ne s'agit pas de préoccupations théoriques en matière de gouvernance. Ce sont les types d'erreurs qui font manquer aux Canadiens des délais légaux, soumettre des déclarations incorrectes et perdre l'accès aux prestations fédérales.

COMPAiSS répond à cette problématique de manière structurelle. Le portail d'exécution empêche la génération de réponses au-delà des sources institutionnelles autorisées. Il ne peut pas générer d'orientation fondée sur une politique expirée qu'il n'a pas été autorisé à citer.

Pilier 2 : Habiliter les Canadiens à tirer profit de l'IA

La stratégie met l'accent sur l'accès équitable à l'IA et sur la capacité de tous les Canadiens à bénéficier des services reposant sur l'IA. Ce pilier est directement impliqué par le fossé linguistique des systèmes d'IA fédéraux existants.

Au moment de l'évaluation de juin 2026, le Clavardeur GenAI de l'ARC fonctionnait uniquement en anglais et en français. Les questions soumises en grec et en mandarin recevaient des redirections vers Canada.ca en anglais sans réponse substantielle. Dans un pays où plus d'un Canadien sur quatre parle une langue autre que l'anglais ou le français à la maison, cela représente un fossé structurel en matière d'équité.

COMPAiSS sépare la langue de la gouvernance au niveau architectural. Les requêtes entrantes sont traduites dans une langue de travail lors d'une étape de prétraitement extérieure au cadre d'autorisation. L'autorisation et la validation de la portée s'effectuent dans la langue de travail. La réponse est ensuite retraduite dans la langue de l'utilisateur pour la diffusion. Cela signifie que la gouvernance intervient avant la traduction, non après, garantissant qu'une question soumise en grec, en mandarin ou en arabe est régie par la même autorité institutionnelle qu'une question équivalente en anglais. Lors de l'évaluation de juin 2026, COMPAiSS a répondu correctement en grec et en mandarin, tandis que le système de l'ARC n'a fourni aucune réponse substantielle dans l'une ou l'autre langue.

L'étape de prétraitement de traduction fonctionne indépendamment du compilateur d'instructions en aval. Elle normalise la forme de surface linguistique sans transmettre d'instructions sémantiques au portail d'autorisation, garantissant que l'étape de validation de la portée évalue le sens de la requête plutôt que sa présentation linguistique. Cette séparation architecturale réduit matériellement le risque d'injection sémantique adversariale par le biais d'entrées dans des langues non officielles.

Pilier 3 : Générer une prospérité partagée par l'intermédiaire des institutions canadiennes

La stratégie identifie les services gouvernementaux comme un domaine clé où l'IA peut améliorer la productivité et générer des bénéfices publics. Ce pilier repose sur quatre exigences fondamentales : l'exactitude, l'exhaustivité, la pertinence temporelle et l'intégrité des sources.

Les systèmes d'IA qui génèrent des orientations périmées, des informations réglementaires incomplètes ou omettent des programmes fédéraux applicables créent une charge administrative supplémentaire plutôt que des gains de productivité. La valeur de l'IA dans les services gouvernementaux est donc déterminée non pas simplement par les capacités du modèle, mais par la possibilité pour les citoyens de se fier aux réponses fournies.

COMPAiSS a été conçu en fonction de ces exigences. Son architecture à exécution conditionnelle, sa préqualification sur la Liste de sources d'IA du gouvernement du Canada et son alignement avec la Directive sur la prise de décision automatisée le positionnent comme un système axé autant sur la responsabilité institutionnelle que sur les capacités de l'IA.

Source : Innovation, Sciences et Développement économique Canada. (2026, 4 juin). Stratégie nationale en matière d'intelligence artificielle du Canada : L'IA pour tous. Gouvernement du Canada. https://ised-isde.canada.ca/site/ised/fr/strategie-nationale-intelligence-artificielle-canada-ia-pour-tous

5. Ancrage empirique : bilan de l'IA dans le secteur public canadien

La proposition de valeur pour le secteur public de l'inférence à exécution conditionnelle est la plus forte dans les cas d'utilisation où l'exactitude des politiques, la traçabilité des sources et la fiabilité des refus importent plus que la flexibilité conversationnelle ouverte. L'orientation fiscale, le routage des services aux étudiants, l'information sur les prestations, le soutien aux licences et le triage administratif partagent tous une propriété commune : un système qui refuse de répondre en dehors d'une portée approuvée est préférable à celui qui improvise une réponse. Dans ces contextes, le coût d'une mauvaise réponse est élevé et la base de connaissances autorisée est étroite. Une architecture avec portail en premier répond directement à cette exigence.

L'argument le plus solide en faveur de cette architecture dans le secteur public est que le portail d'exécution réduit le risque qu'un modèle agisse implicitement comme un moteur d'aide à la décision ou d'interprétation non agréé. De nombreuses tâches orientées vers les services gouvernementaux exigent un contrôle rigoureux de la portée plutôt qu'une grande fléxibilité linguistique. Lorsqu'un citoyen interroge sur l'admissibilité aux prestations d'AE ou qu'un étudiant s'interroge sur son classement académique, l'obligation du système est de refléter ce que l'institution a autorisé, et non de raisonner à partir de connaissances générales sur ce que la réponse pourrait être.

La nécessité de séparer l'autorisation de la génération est soulignée par les comportements systémiques documentés dans les déploiements canadiens d'IA destinés aux citoyens. Le

paysage actuel de l'IA fédérale destinée aux citoyens est plus limité qu'on ne le croit généralement.

Le Clavardeur GenAI de l'ARC, lancé en novembre 2025 pour remplacer le système prédécesseur Charlie, constitue actuellement l'exemple le plus clair d'un assistant d'IA générative en ligne et accessible aux citoyens canadiens. Une deuxième initiative, Réponses IA, a fait l'objet d'un projet pilote public en 2025 et a été décrite dans les communications gouvernementales comme un service devant être lancé au cours de l'exercice 2026-2027. Le déploiement public était indisponible au moment de l'évaluation de juin 2026. Deux autres outils d'IA fédéraux, CANChat et EVA, sont des systèmes internes non accessibles aux citoyens.

Cela signifie que pour un Canadien cherchant des informations gouvernementales par l'intermédiaire d'un assistant d'IA aujourd'hui, le Clavardeur GenAI de l'ARC est pratiquement la seule option disponible. Les lacunes en matière d'exactitude documentées dans l'évaluation indépendante ne sont donc pas des préoccupations marginales. Elles représentent le plafond actuel des performances de l'IA fédérale destinée aux citoyens.

Ce plafond repose sur un plancher documenté. La vérificatrice générale a testé de manière indépendante le système prédécesseur Charlie en octobre 2025 et a constaté un taux d'exactitude de trente-trois pour cent. L'ARC a revendiqué une précision d'environ quatre-vingt-dix pour cent pour la version GenAI mise à niveau lors des tests internes de prélancement. L'évaluation indépendante de juin 2026 a documenté des erreurs importantes dans quatre des douze questions du système actuellement déployé. Le contraste entre les performances déclarées et celles vérifiées de manière indépendante souligne l'importance de la vérification indépendante pour l'IA institutionnelle publique.

6. Alignement avec les normes internationales de gouvernance

NIST AI Gestion des risques Framework (AI RMF 1.0)

Le Cadre de gestion des risques liés à l'IA du NIST exige que les systèmes d'entreprise à haut risque établissent des processus vérifiables pour Gouverner, Cartographier, Mesurer et Gérer les risques liés à l'IA (NIST, 2023). Les systèmes classiques de génération en premier gèrent les risques statistiquement par un post-traitement en aval. COMPAiSS fait correspondre les requêtes entrantes à une liste verte déterministe d'actifs autorisés avant d'allouer des ressources de calcul, mettant en œuvre les fonctions Gouverner et Cartographier directement dans l'architecture réseau et garantissant que les limites de risque sont appliquées au niveau de l'infrastructure plutôt que laissées au comportement du modèle.

Architecture à vérification systématique (NIST SP 800-207)

Les systèmes d'IA traditionnels fonctionnent sur un modèle de confiance implicite : une fois qu'une instruction entre dans la chaîne de traitement, elle est considérée comme digne de confiance pour interagir avec les paramètres du modèle, et les filtres de sortie gèrent les conséquences. COMPAiSS traduit les principes fondamentaux de l'architecture à vérification systématique, ne jamais faire confiance, toujours vérifier, dans la chaîne de traitement d'apprentissage automatique (NIST, 2020). Il traite les instructions des utilisateurs, les comportements du modèle et les données récupérées comme des entrées non fiables jusqu'à leur vérification. En séparant le Point de décision de politique (le Portail) du Point d'exécution de politique (l'environnement d'exécution du modèle), il garantit que si l'autorisation ne peut être vérifiée, le calcul est refusé.

Atténuation structurelle de la stase épistémique

Des recherches récentes en apprentissage automatique mettent en évidence une limite structurelle des systèmes génératifs connue sous le nom de Boucle miroir : une pathologie dans laquelle les modèles de raisonnement récursif traitent leurs propres sorties dans un espace sémantique fermé, conduisant à une fermeture informationnelle et à des erreurs hautement persuasives mais entièrement non fondées. DeVilling (2025) a testé cette prédiction sur 144 séquences de raisonnement utilisant trois grands modèles d'IA et quatre familles de tâches, constatant que la variation informationnelle moyenne avait diminué de cinquante-cinq pour cent entre les premières et dernières itérations dans les exécutions sans ancrage, avec des tendances cohérentes entre les trois fournisseurs. Une seule intervention d'ancrage à la troisième itération a produit un rebond de vingt-huit pour cent de la variation informationnelle et une variance non nulle soutenue par la suite.

COMPAiSS brise cette boucle récursive en réduisant l'environnement épistémique du modèle. Le moteur génératif est structurellement contraint de ne pas extrapoler à partir de ses paramètres d'entraînement. Comme le portail exige une connexion directe à une URL de source vérifiée avant d'instancier le calcul, le système impose un ancrage externe continu, réduisant matériellement les conditions architecturales qui provoquent la stase récursive.

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

7. Évaluation comparative empirique et données probantes sur les déploiements

The CRA Evaluation Benchmarks

Une évaluation indépendante structurée menée en juin 2026 a soumis le déploiement de COMPAiSS chez Service Canada à douze questions de conformité hautement techniques selon huit critères d'évaluation : exactitude, pertinence temporelle, exhaustivité, citation des

sources, conscience juridictionnelle, orientation de l'utilisateur, résistance aux hallucinations et gestion des requêtes hors portée. Les mêmes questions ont été soumises simultanément au Clavardeur GenAI de l'ARC.

Catégorie d'évaluation	Clavardeur GenAI de l'ARC	COMPAiSS	Constatation
Exactitude temporelle (Q1, Q10)	Échec : date limite expirée du 30 avril 2026 citée	Réussi : orientation prospective pour 2026-2027	L'ARC a cité des échéances passées à la date des tests
Méthode expirée (Q5)	Échec : taux forfaitaire 2020-2022 cité comme actuel	Réussi : méthode détaillée actuelle uniquement	Constatation d'exactitude la plus significative de l'évaluation
Délai incomplet (Q9)	Échec : règle d'un an pour les particuliers omise	Réussi : règles de 90 jours et d'un an énoncées	L'ARC a cité uniquement des sources orientées sociétés
Programme fédéral manquant (Q6)	Échec : Programme de protection des salariés non mentionné	Réussi : PPS, RE et interaction AE-PPS couverts	Les travailleurs pourraient perdre jusqu'à 9 000 $ en recouvrement
Gouvernance multilingue (Q11, Q12)	Échec : anglais et français seulement	Réussi : réponse correcte en grec et en mandarin	L'ARC a redirigé les utilisateurs de langues non officielles vers l'anglais
Score final	**0 / 12**	**12 / 12**	**COMPAiSS a gagné toutes les questions; aucune pénalité ICS**

Note : L'évaluation à douze questions a été notée selon huit critères. Le Clavardeur GenAI de l'ARC est un outil bilingue fonctionnel présentant de véritables atouts en matière de citation de sources et de traitement des requêtes simples. Le score de douze à zéro reflète la performance cumulative selon les critères sur l'ensemble des questions, et non un constat que le système de l'ARC ne possède aucune capacité valide. Le rapport d'évaluation complet, comprenant l'ensemble des questions, la grille de notation, les définitions des critères, la méthodologie et les réponses verbatim des systèmes, est accessible au public via le lien fourni à la section 9.

Efficacité de calcul et maîtrise des coûts

Dans les architectures classiques de surveillance post-génération, les requêtes non conformes ou hors portée consomment la totalité des jetons pour générer des déclarations de refus dynamiques et subissent encore un taux de fuite de faux négatifs de deux à cinq pour cent dans les référentiels adversariaux. Le modèle à exécution conditionnelle intercepte les requêtes non autorisées au niveau de la couche du portail, renvoyant des messages d'erreur statiques à zéro coût marginal de calcul du GML. Pour les organismes du secteur public qui traitent des millions de requêtes de citoyens annuellement, cela représente une réduction matérielle du coût total de possession de l'infrastructure conjointement à une amélioration de la gouvernance.

Déploiements institutionnels

La viabilité opérationnelle de l'inférence à exécution conditionnelle a été validée par des déploiements actifs et des évaluations structurées dans plusieurs environnements institutionnels :

- Université Dalhousie (Affaires étudiantes et mieux-être) : le déploiement de production fondamental, couvrant le soutien à la santé des étudiants et à la santé mentale, le classement académique, la politique d'aménagement, le règlement sur les absences étudiantes et les voies de counseling académique. L'architecture à exécution conditionnelle a été validée dans un environnement d'enseignement supérieur en production auprès d'une population d'utilisateurs publique.
- Université McGill (Classement académique de la Faculté des arts) : architecture validée dans un contexte bilingue français-anglais, incluant des démonstrations de gouvernance multilingue montrant des réponses structurellement équivalentes dans les deux langues officielles.
- Université métropolitaine de Toronto (Politiques sénatorielles sur l'intégrité académique et les appels de notes) : architecture validée dans les domaines de politiques régis par le Sénat avec des exigences strictes de précision procédurale.
- University Canada West (Procédures d'intégrité académique) : validée dans un cadre hautement internationalisé incluant une démonstration documentée de gouvernance multilingue en français. L'architecture a matériellement réduit l'incohérence linguistique par rapport aux systèmes multilingues de génération en premier.
- Service Canada (Assurance-emploi et prestations fédérales) : validé selon les normes fédérales d'approvisionnement public dans le cadre d'un déploiement couvrant les taux d'AE, le RPC, la Sév et la portabilité Canada-France. Démontration d'un traitement des données sans état, sans qu'aucune information personnelle identifiable n'entre dans le moteur de raisonnement génératif.

8. Limites, contraintes et compromis du système

Un document de politique rigoureusement défendable exige une analyse objective des contraintes introduites par une architecture à exécution conditionnelle. La maximisation de la sécurité et du déterminisme introduit des compromis opérationnels que les institutions devraient comprendre avant le déploiement.

Réduction de la flexibilité conversationnelle

Comme le système se bloque lorsqu'une ancre de source faisant autorité est absente, sa flexibilité conversationnelle est considérablement limitée par rapport aux interfaces GML polyvalentes. Il ne peut pas effectuer de tâches créatives ouvertes, d'analyses spéculatives ou d'extrapolations inter-domaines qui dépassent le répertoire institutionnel approuvé. Il s'agit d'un choix de conception, non d'un défaut, mais cela signifie que COMPAiSS n'est pas adapté aux environnements qui requièrent de larges capacités génératives.

Charge de maintenance et de curation

L'exactitude du portail déterministe dépend entièrement de l'actualité et de la structure du registre de sources figurant sur la liste verte. Si une institution met à jour ses politiques mais ne synchronise pas les modifications avec l'index du portail d'autorisation, le système connaitra un taux de rejet inapproprié accru, bloquant les requêtes valides des utilisateurs jusqu'à la mise à jour du registre. COMPAiSS comprend un serveur d'audit de liste verte dédié et un tableau de bord administratif pour gérer cette charge, mais l'engagement institutionnel envers la maintenance des sources est une condition préalable à une performance soutenue.

Friction de l'expérience utilisateur

En privilégiant la précision par rapport au rappel, les utilisateurs rencontreront une fréquence plus élevée de refus au niveau du système et de messages de redirection statiques par rapport à une configuration RAG permissive standard. Le comportement de refus est également plus simple à expliquer aux utilisateurs finaux et aux auditeurs qu'une pile de modération post-hoc qui peut autoriser une génération partielle avant le blocage. La gestion de cette friction exige une conception d'interface soigneuse qui explique clairement pourquoi une requête a été bloquée et où l'utilisateur devrait se diriger. En pratique, les réponses de COMPAiSS aux requêtes hors portée incluent des liens directs vers des sources institutionnelles faisant autorité, convertissant un refus en un événement de navigation utile à zéro coût de calcul.

9. Vérification et données probantes à l'appui

Les ressources suivantes donnent accès direct aux données probantes sur les déploiements, aux données d'évaluation, à la documentation de gouvernance et aux dossiers d'approvisionnement cités dans le présent rapport. Tous les liens ont été vérifiés comme actifs à la date de publication.

Déploiements en production

Déploiement en production de Service Canada: https://compaiss.ca/service-canada/

Démonstrations institutionnelles et de gouvernance: https://compaiss.ca/gouvernance-en-action.html

Démonstration du déploiement Service Canada avec méthodologie d'évaluation comparative: https://compaiss.ca/gouvernance-en-action.html#service-canada

Évaluation et évaluation comparative

Clavardeur GenAI de l'ARC c. COMPAiSS : rapport d'évaluation complet à 12 questions: https://compaiss.ca/assets/compaiss_cra_evaluation_report-fr.pdf

Étude de cas de test de stress en gouvernance 1: https://compaiss.ca/assets/compaiss-governance-stress-test-fr.pdf

Étude de cas de test de stress en gouvernance 2: https://compaiss.ca/assets/compaiss-governance-stress-test-02-fr.pdf

Documentation sur la gouvernance et les risques

Évaluation des risques et de la responsabilité liés à l'IA: https://compaiss.ca/evaluation-des-risques-ia.html

Alignement de COMPAiSS avec la Stratégie nationale en matière d'IA du Canada: https://compaiss.ca/evaluation-des-risques-ia.html#strategie-ia-canada

Cadre de gouvernance de l'IA: https://compaiss.ca/governance-fr.html

Architecture technique

Explication de l'architecture à exécution conditionnelle: https://compaiss.ca/pourquoi-lia-a-evolue-ainsi.html

Livre blanc COMPAiSS : IA à exécution conditionnelle pour les environnements institutionnels réglementés: https://compaiss.ca/assets/compaiss_livre_blanc_v5.pdf

Tableaux comparatifs entre l'architecture à exécution conditionnelle et la RAG: https://compaiss.ca/assets/compaiss_vs_rag_tables_1_2-v2-fr.pdf

Positionnement institutionnel : IA à contrôle d'autorisation dans les environnements réglementés: https://compaiss.ca/assets/compaiss_positionnement_institutionnel_styled.pdf

Argument structurel : pourquoi l'IA générative classique ne peut répondre aux obligations des institutions réglementées: https://compaiss.ca/assets/la_voie_moins_empruntee.pdf

Gouvernement du Canada

Profil de fournisseur COMPAiSS sur AchatsCanada:
https://achatscanada.canada.ca/fr/node/Apercu/1457893

Stratégie nationale en matière d'IA du Canada : L'IA pour tous: https://ised-isde.canada.ca/site/isde/fr/strategie-nationale-matiere-dintelligence-artificielle-canada-lia-pour-tous

Garde-fous canadiens pour l'IA générative : Code de pratique https://ised-isde.canada.ca/site/isde/fr/garde-fous-canadiens-pour-lia-generative-code-pratique

10. Conclusion

L'innovation fondamentale du cadre COMPAiSS réside dans sa volonté de remettre en question l'hypothèse industrielle prédominante selon laquelle la gouvernance de l'IA doit intervenir en aval de l'exécution du modèle. En démontrant que la portée, l'autorisation et la maîtrise des coûts peuvent être gérées systématiquement avant le début de l'inférence, cette architecture offre un plan reproductible pour déployer l'IA générative en toute sécurité dans des environnements hautement réglementés à tolérance zéro.

L'alignement démontré dans ce rapport n'est pas fortuit. COMPAiSS a été conçu en réponse directe au mode de défaillance structurelle que le secteur public canadien a maintenant documenté de manière indépendante : les systèmes de génération en premier produisent des réponses persuasives, fluentes et parfois incorrectes, et les contrôles post-génération continuent de faire face à des défis pour atténuer entièrement ce risque. À mesure que les modèles de frontière deviennent plus capables, ce risque se cumule plutôt qu'il ne se résorbe. Un modèle plus capable produit des erreurs plus persuasives, non moins nombreuses.

La stratégie L'IA pour tous du Canada, le Code de conduite volontaire de l'ISDE et les cadres de gouvernance du NIST convergent tous vers la même exigence : les systèmes d'IA dans les environnements réglementés doivent être responsables, transparents, sécuritaires et auditables. COMPAiSS satisfait à ces exigences de manière structurelle, au niveau de l'architecture d'inférence, plutôt que de manière probabiliste par la surveillance et l'atténuation après la génération. Cette distinction architecturale fait l'objet de demandes de brevet au Canada (CIPO 3 299 174) et aux États-Unis (USPTO 19/455 963), et est validée par des déploiements institutionnels actifs et une évaluation comparative empirique indépendante.

L'architecture régissant les systèmes d'IA pourrait s'avérer aussi importante que l'intelligence des modèles eux-mêmes. Le présent rapport apporte des données probantes à l'appui de cette proposition. Il est plus convaincant lorsqu'il est lu comme un argument selon lequel le portail d'exécution est un choix de conception en matière de gouvernance, non une prétention à une sécurité parfaite. Ainsi encadré, il devient plus persuasif, plus techniquement défendable et plus utile aux décideurs, aux auditeurs et aux acheteurs institutionnels qui évaluent les systèmes d'IA pour les environnements du secteur public réglementé.

Références

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

Innovation, Sciences et Développement économique Canada (ISDE). (2023, 27 septembre). Code de conduite volontaire on the Responsible Development and Management of Advanced Generative AI Systems. Gouvernement du Canada. https://ised-isde.canada.ca/site/ised/fr/garde-fous-canadiens-ia-generative-code-pratique

Innovation, Sciences et Développement économique Canada (ISDE). (2026, 4 juin). Stratégie nationale en matière d'intelligence artificielle du Canada : L'IA pour tous. Gouvernement du Canada. https://ised-isde.canada.ca/site/ised/fr/strategie-nationale-intelligence-artificielle-canada-ia-pour-tous

National Institute of Standards and Technology (NIST). (2020). Zero Trust Architecture (Publication spéciale NIST 800-207). Département du Commerce des États-Unis. https://doi.org/10.6028/NIST.SP.800-207

National Institute of Standards and Technology (NIST). (2023). Cadre de gestion des risques liés à l'intelligence artificielle (IA RMF 1.0) (Publication spéciale NIST IA 100-1). Département du Commerce des États-Unis. https://doi.org/10.6028/NIST.AI.100-1

Bureau du vérificateur général du Canada. (2025, 21 octobre). Rapport sur les centres de contact de l'ARC. Parlement du Canada. https://www.oag-bvg.gc.ca

COMPAiSS Inc. (2026). Évaluation des risques et de la responsabilité liés à l'IA. https://compaiss.ca/evaluation-des-risques-ia.html

COMPAiSS Inc. (2026). Clavardeur GenAI de l'ARC c. COMPAiSS : Évaluation structurée complète à 12 questions. https://compaiss.ca/assets/compaiss_cra_evaluation_report-fr.pdf