



Gouverner l'IA : Un plan pour le Canada

Cinq actions à mener en
2023/2024

Un livre blanc de

AIGS  **GSIA**

Octobre 2023



Gouvernance et Sécurité de l'IA Canada est une organisation à but non lucratif multipartite et une communauté de personnes à travers le pays qui travaillent à renforcer le leadership du Canada en matière de gouvernance et de sécurité de l'intelligence artificielle (IA). Notre point de départ est la question "Que pouvons-nous faire au Canada, et à partir du Canada, pour s'assurer que l'avenir créé par l'IA soit bénéfique?"

L'objectif de ce document est d'informer les décideurs politiques sur les moyens nécessaires pour garantir la sécurité des technologies d'IA pour les Canadiens et l'humanité en général. Il offre une vue d'ensemble de l'évolution des risques, aborde les principaux enjeux de la gouvernance de l'IA, et se conclut avec cinq mesures que le gouvernement canadien peut prendre dès aujourd'hui.

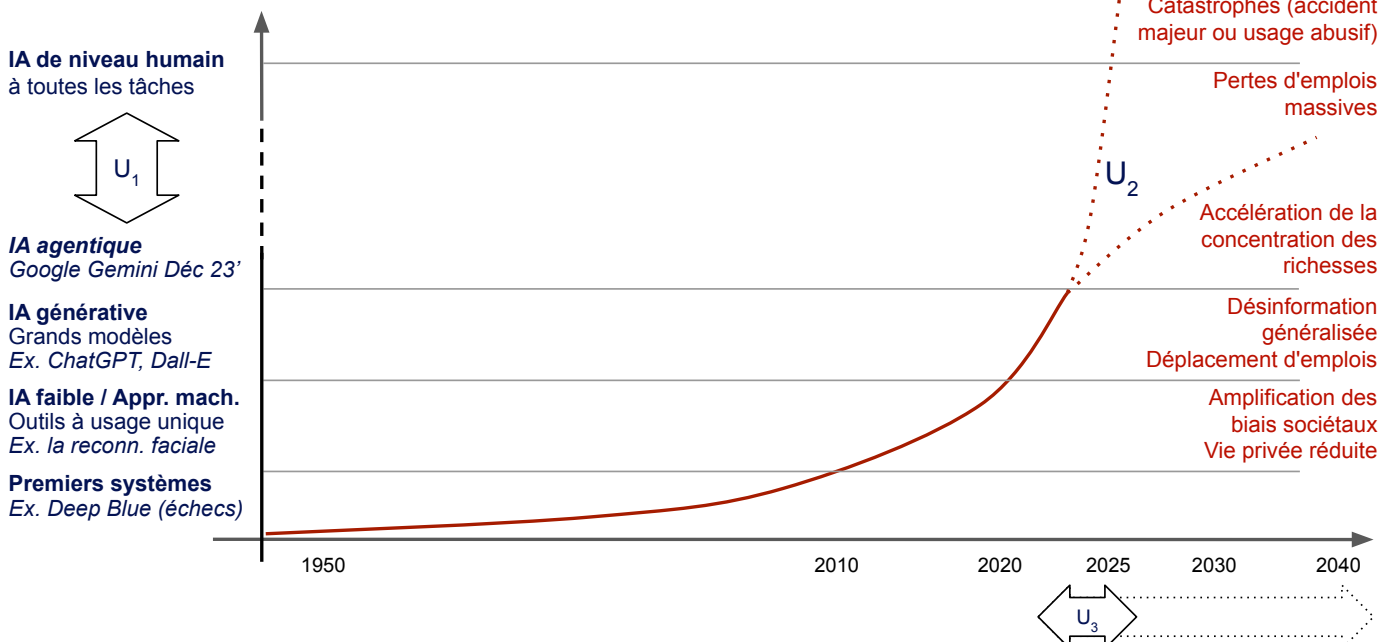
Auteurs : Wyatt Tessari L'Allié, Gabrielle Tran

Avec le soutien des membres de GSIA et de son conseil d'experts.

L'évolution des risques liés à l'IA

Nous témoignons actuellement d'une révolution dans le domaine de l'IA, où les capacités croissantes entraînent des avantages et des risques croissants. Les années 2010 ont vu l'apparition d'algorithmes à usage unique très utiles, capables de décisions d'embauche à la reconnaissance faciale, en passant par les armes autonomes létales. Malheureusement, dû aux limites de certaines techniques et aux biais des données humaines utilisées, plusieurs personnes se sont vues injustement refuser un emploi, un prêt hypothécaire ou une caution. Les années 2020 ont ensuite donné le coup d'envoi aux grands modèles tels que ChatGPT et Midjourney, capables d'écrire des textes intelligents et de générer des images de haute qualité. Avec ces percées, les cybercriminels ont de nouveaux outils, les industries créatives se retrouvent perturbées, et les deepfakes et la désinformation de masse mettent en péril le discours public et la démocratie. L'IA aggrave également les inquiétudes préexistantes concernant la protection de la vie privée, la concentration des richesses dans les entreprises technologiques, et la fracture numérique qui laisse de nombreuses personnes à l'écart.

Capacités de l'IA



Graphique 1 : Esquisse des capacités et des impacts de l'IA au fil du temps

Les principaux domaines d'incertitude sont U₁ : le nombre de percées techniques nécessaires pour atteindre l'IA de niveau humain et U₂ : la poursuite du rythme exponentiel actuel des progrès. Cela détermine U₃ : le temps dont disposent les gouvernements pour se préparer aux risques catastrophiques potentiels et aux bouleversements sociaux majeurs.

Bien que ces deux premières vagues aient été perturbatrices, il est probable que d'autres, encore plus importantes, sont à venir. Les capacités de l'IA sont en passe de surpasser celles des humains dans toutes les tâches, y compris la stratégie, l'acquisition de ressources, l'interaction humaine, les découvertes scientifiques et l'amélioration de leur propre intelligence.

L'IA de niveau humain ouvre la perspective de l'automatisation du développement technologique, de pertes d'emplois massives et de nouveaux risques. Celles-ci incluent notamment les catastrophes mondiales dues à une mauvaise conception, à une utilisation malveillante ou à un accident. Les experts ne s'accordent pas sur le délai dans lequel l'IA de niveau humain sera construite les estimations allant de 2025 à 2060. Toutefois, avec les capacités des modèles récents dépassant largement les attentes, les marchés de prédiction se sont considérablement orientés vers des échéances plus courtes. Le graphique 1 et le tableau 1 résument cette discussion sur les capacités de l'IA et leurs impacts au fil du temps, ainsi que les principales incertitudes pour l'avenir.

Catégorie d'IA	IA faible Apprentissage machine à usage unique	Grands modèles Systèmes d'IA à usage général IA générative	Vers l'IA de niveau humain Et au-delà
Chronologie	Depuis ~2010	Depuis ~2020	2025 ? 2040 ? (inconnu)
Capacités principales	Reconnaissance faciale Traitement du langage naturel Reconnaissance vocale Recommandations	Capacités d'IA faible, et : Génération de texte, d'audio et d'images Synthèse de l'information Séquençage des gènes Développement logiciel	Capacités des grands modèles, et : Conscience avancée de sa situation Planification et action à long terme Acquisition de ressources Ingénierie sociale Auto-amélioration autonome
Principaux impacts à prendre en compte	Biais algorithmique Surveillance Transparence Chambres d'écho Empreinte carbone Armes autonomes létales	Impacts de l'IA faible, et : Perturbation des industries créatives Déplacement d'emplois accéléré Maltraitance des étiqueteurs de données à l'étranger Désinformation généralisée Fracture numérique / inégalité Compétition accélérée	Préoccupations grand modèles, et : Perte massive d'emplois Concentration extrême des richesses Désordre social dû au changement rapide Systèmes IA trompeurs Dynamique militaire déstabilisée Accidents majeurs entraînant une catastrophe

Tableau 1 : Détail des capacités et des impacts de l'IA par catégorie

Les colonnes détaillent les trois grandes catégories d'IA énumérées dans le graphique 1, qui représentent les capacités récentes, actuelles et potentielles à l'avenir, ainsi que les préoccupations correspondantes.

Considérations pour la gouvernance de l'IA

Tous les gouvernements sont confrontés à des défis majeurs dans la gestion de l'IA, tels que la rapidité de son développement, les difficultés à réglementer les produits non physiques et les incitations économiques et géopolitiques actuelles.

L'IA est appelée à jouer un rôle croissant dans l'[économie](#) et les [forces armées](#), et les laboratoires et pays développant l'IA sont engagés dans une course ouverte pour être les premiers à déployer des modèles de pointe, ce qui met en péril l'éthique et la sécurité. Tout pays qui tenterait de réglementer seul l'IA étoufferait l'innovation sur son territoire et pousserait les laboratoires vers d'autres juridictions, qui pourraient en tirer un avantage décisif. Une gouvernance efficace de l'IA est donc impossible sans une forte collaboration mondiale.

Le Canada a déjà pris un certain nombre d'initiatives avec sa [stratégie pancanadienne en matière d'IA](#) de 2017 (SPCIA), sa [directive sur la prise de décision automatisée](#) (2019) et l'introduction du [projet de loi C-27 incluant la Loi sur l'IA et les données \(LIAD\)](#) en juin 2022, et, plus récemment, pour l'IA générative, un [guide pour le secteur publique](#) et une [code de pratique volontaire](#) pour le secteur privé.

Bien que ces efforts soient louables, la majorité des investissements du SPCIA ont servi à augmenter les capacités de l'IA, les directives ne s'appliquent qu'au gouvernement et le code de conduite n'est pas contraignant. S'il est adopté, le projet de loi C-27 introduira des règles strictes ainsi que des responsabilités, mais il comporte un certain nombre de limites. Il s'agit notamment du manque de clarté quant aux systèmes d'IA auxquels il s'applique, du manque de capacités et d'indépendance du commissaire à l'IA et d'une date d'application repoussée au moins jusqu'en 2025. Le plus préoccupant est qu'aucun de ces efforts ne vise à préparer le pays aux risques encore plus importants à venir (tableau 1).

Il est donc urgent d'adopter une approche plus solide. Cela permettra non seulement de mieux protéger les Canadiens, mais aussi de bâtir le leadership du Canada sur ce qui pourrait devenir l'enjeu déterminant du 21e siècle.

Un plan pour le Canada : Cinq actions à mener en 2023/24

Compte tenu du rythme de développement de l'IA et de l'ampleur des impacts potentiels, nous recommandons au Canada de prendre les cinq mesures suivantes pour faire progresser la gouvernance de l'IA **d'ici l'été 2024** :

1 **Créer une agence centrale de l'IA pour le gouvernement du Canada**

- ◆ Basé au sein du Bureau du Conseil privé
- ◆ Suit les développements en matière d'IA, informe les dirigeants, communique avec les provinces et coordonne les actions du gouvernement fédéral

Raisonnement: Dans un environnement qui évolue rapidement, le gouvernement a besoin de capacités internes pour être efficace. Une équipe d'experts au sein d'une agence centrale sera en mesure de garantir que le CPM et les ministères disposent des informations dont ils ont besoin pour limiter les risques et exploiter les opportunités au fur et à mesure qu'elles se présentent. L'ISDE ou d'autres départements n'ont pas les incitations institutionnelles ou le mandat pour naviguer l'équilibre entre tous les enjeux.

2 Investir > 500 millions de dollars dans la recherche sur la sécurité et la gouvernance de l'IA

- ◆ Plus que doubler le financement actuel de la stratégie pancanadienne en matière d'IA gérée par le CIFAR
- ◆ Diriger les fonds supplémentaires vers l'amélioration des politiques de gouvernance de l'IA et des solutions techniques de sécurité

3 Soutenir et financer les efforts de gouvernance de l'IA au niveau mondial

- ◆ Appuyer les initiatives de gouvernance tels qu'à l'ONU, au G7, au [sommet britannique](#), et à l'OCDE
- ◆ Mener la conversation sur les [modèles du CERN, du GIEC et de l'AIEA](#) pour le développement de l'IA avancée

4 Améliorer et adopter la loi sur l'IA et les données

- ◆ Meilleur mécanisme pour l'adoption de politiques en matière d'audit, de transparence et de responsabilité légale
- ◆ Créer une commission de l'IA, chargée de réglementer les modèles à haut risque

5 Lancer un dialogue national sur l'IA

- ◆ Organiser des consultations publiques ouvertes sur l'IA afin d'éclairer les décisions politiques sur la transition des emplois, la concentration des richesses et le juste équilibre entre les bénéfices et les risques

Raisonnement: Le monde ne dispose actuellement ni des solutions techniques ni des mécanismes politiques efficaces pour garantir une IA sûre et éthique. Avec des centaines de milliards investis chaque année pour renforcer les capacités et des délais très incertains pour les formes les plus dangereuses d'IA, les gouvernements doivent jouer un rôle majeur pour s'assurer que la gouvernance et la recherche en matière de sécurité rattrapent rapidement leur retard.

Raisonnement: En l'absence de coordination mondiale, la réglementation nationale ne fera que pousser le développement de l'IA vers d'autres juridictions. De plus, il est probable que les premiers systèmes de niveau humain soient construits en dehors du Canada, ce qui signifie que pour réduire les risques catastrophiques il faudra aussi influencer les actions d'autres pays (en particulier les États-Unis et la Chine). Pour faire avancer les négociations, le monde a besoin de pays leaders. C'est l'opportunité pour le Canada de jouer un rôle de premier plan.

Raisonnement: L'industrie et le public ont besoin d'un cadre réglementaire qui clarifie ce qui constitue un développement et une utilisation acceptables de l'IA. Il est donc urgent d'accélérer les consultations et la rédaction de la législation. Bien qu'une élection anticipée puisse faire échouer le projet de loi C-27, tout progrès réalisé sur l'actuel LIAD fera avancer la conversation et sera utile aux futurs gouvernements.

[Les détails de nos recommandations sur la LIAD sont disponibles à GSIA.ca](#)

Raisonnement: Même si les risques pour la sécurité sont gérés, l'IA transformera radicalement la société, créant potentiellement un monde post-emploi et post-pénurie dans lequel l'intelligence humaine et les institutions seront reléguées au second plan. Gérer cette transition et s'assurer qu'elle reflète la volonté des citoyens prendra du temps et doit commencer tôt. Grâce à sa stabilité relative, à sa population éduquée et à son écosystème de l'IA, le Canada est particulièrement bien placé pour piloter une telle conversation.