

Gouverner l'IA: Un plan pour le Canada

Cinq actions à mener en
2024/2025

Un livre blanc de



Juin 2024

Gouvernance et Sécurité de l'IA Canada (GSIA Canada) est un organisme à but non lucratif et non partisan, ainsi qu'une communauté de personnes à travers le pays, œuvrant pour que l'IA avancée soit sécuritaire et bénéfique pour tous.

L'objectif de ce document est d'informer les décideurs politiques et le public 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.

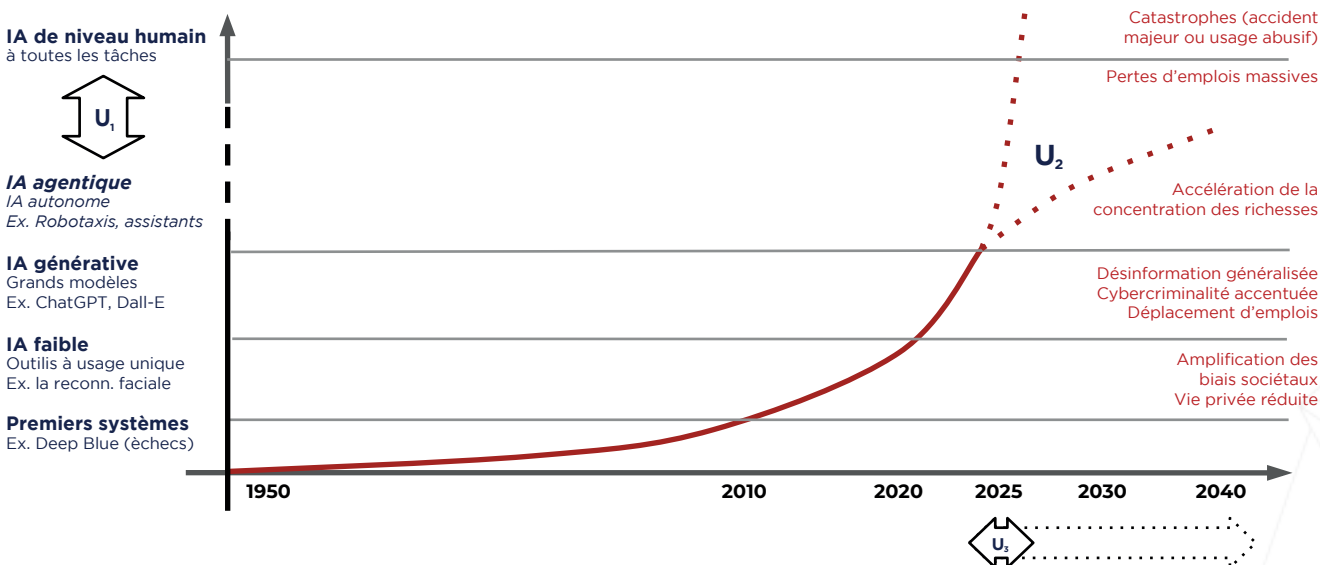
Auteurs: Wyatt Tessari L'Allié (Fondateur, GSIA Canada) et Kathrin Gardhouse (Responsable des politiques, GSIA Canada), avec le soutien et l'expertise des membres et du conseil consultatif de GSIA.

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.

Bien que ces deux premières vagues aient été perturbatrices, il est probable que d'autres, encore plus importantes, suivront. 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 dont notamment les [catastrophes mondiales](#) dues à [une utilisation malveillante ou une perte de contrôle](#). 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](#). Cependant, son potentiel économique et sa faisabilité à court terme poussent de nombreuses entreprises – y compris [OpenAI](#), [Google DeepMind](#), [Meta](#) et [Salesforce](#) – à faire de l'IA de niveau humain leur objectif principal. Le graphique 1 et le tableau 1 résument cette discussion.

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 : le rythme des progrès. Cela détermine U₃ : le temps dont disposent les gouvernements pour se préparer aux risques catastrophiques potentiels et aux bouleversements sociaux majeurs.

Catégorie d'IA	IA faible Apprentissage machine à usage unique	IA générative Grands modèles Systèmes d'IA à usage général	Vers l'IA de niveau humain Et au-delà
Chronologie	Depuis ~2010	Depuis ~2020	2025 ? 2040 ? (inconnu)
Capacités principales	Reconnaissance faciale et vocale Traitement du langage naturel Recommandations	Capacités d'IA faible, et: Génération texte, audio, images Synthèse de l'information Développement logiciel Influence et persuasion	Capacités des grands modèles, et: Conscience avancée de sa situation Planification et action à long terme Acquisition de ressources Raisonnement et sens commun Auto-amélioration autonome
Principaux impacts à prendre en compte	Biais algorithmiques Surveillance Transparence Chambres d'écho Empreinte carbone Armes autonomes létales	Impacts de l'IA faible, et: Industries créatives perturbées Déplacement d'emplois accéléré Désinformation généralisée Hypertrucage pornographique Fracture numérique / inégalité Course au développement	Préoccupations grand modèles, et: Perte massive d'emplois Concentration extrême des richesses Désordre social 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.

La réponse jusqu'à présent

À l'échelle mondiale, les mesures visant à remédier aux effets néfastes de l'IA varient d'incomplètes à inexistantes. Les problèmes de biais sont les mieux explorés, avec des [recherches](#) approfondies, de nouvelles lois proposées et une industrie croissante de [solutions](#), sans que ce soit suffisant. Les questions de droits d'auteur et de responsabilité sont partiellement couvertes par les lois existantes, mais beaucoup de choses restent à [décider](#). La cybercriminalité assistée par l'IA est un phénomène en pleine expansion qui exacerbe les vulnérabilités existantes dans l'infrastructure numérique et qui n'est que partiellement compensée par la [cybersécurité](#) assistée par l'IA. En ce qui concerne la désinformation et les "deepfakes", un certain nombre d'approches telles que [l'interdiction](#) et le [filigrane](#) ont été suggérées, mais les solutions techniques et les mécanismes de contrôle sont au mieux balbutiants. Les [déplacements d'emplois](#) dus à l'IA générative se poursuivent et, si la vitesse et l'ampleur des perturbations continuent de s'accroître, les programmes de reconversion traditionnels seront probablement insuffisants. Et le plus préoccupant est le manque de prévoyance et de planification de la part des gouvernements et des technologues concernant les implications économiques, politiques et de sécurité publique de l'IA de niveau humain, ainsi que le manque de sensibilisation et de participation du public.

Parmi les efforts notables en matière de gouvernance de l'IA, on peut citer la [Loi sur l'intelligence artificielle de l'Union européenne](#), qui aborde une série d'impacts allant des cas d'utilisation à haut risque à la surveillance des modèles à usage général. Les États-Unis ont publié un [décret](#) définissant des lignes directrices sur la sécurité des modèles d'avant-garde et envisagent à présent une action [bipartite](#) au sein du Congrès. Pour sa part, le Royaume-Uni a lancé le premier institut de sécurité de l'IA au monde et a inspiré une série de [rapports scientifiques](#) sur la sécurité de l'IA avancée, inspirée par le modèle [GIEC](#).

Le Canada a déjà pris un certain nombre d'initiatives avec sa [stratégie pancanadienne](#) en matière d'IA (2017), sa [directive sur la prise de décision automatisée](#) (2019), sa cofondation du [Partenariat mondial sur l'IA](#) en 2020 et la présentation du projet de [Loi sur l'IA et les données](#) (LIAD) en juin 2022. En 2023, Ottawa a mis en place un [guide](#) gouvernemental et un [code de conduite volontaire](#) pour l'IA générative, et a signé la [Déclaration de Bletchley](#). En 2024, il a annoncé des [investissements budgétaires](#) de 2,4 milliards de dollars, pour donner aux organisations canadiennes plus d'accès au calcul, et pour la création d'un Institut de sécurité de l'IA (ISIA) qu'il s'est engagé à [coordonner](#) avec des institutions partenaires à l'échelle mondiale.

Bien que ces efforts soient louables, il reste encore beaucoup à faire. Les capacités et les structures du gouvernement sont insuffisantes pour faire face à la croissance des risques liés à l'IA. La [LIAD](#) est toujours en comité et comporte des [limitations](#) importantes qui devront être résolues. Les 50 millions de dollars de financement de l'ISIA sont insuffisants compte tenu de l'ampleur des besoins en matière de sécurité de l'IA, et bien que le Canada ait approuvé des résolutions clés au [G7](#), à [l'ONU](#) et sur l'utilisation [militaire](#), ces dernières années, il a surtout été un spectateur. Entre-temps, le public comprend mal les avantages et les risques de l'IA, et sa confiance dans l'IA est [l'une des plus faibles au monde](#). Il est donc urgent d'adopter une vision claire des besoins et de mettre en œuvre des mesures concrètes.

Considérations relatives à la gouvernance de l'IA

Les gouvernements sont confrontés à plusieurs défis majeurs dans le domaine de l'IA. Tout d'abord, le rythme de développement depuis 2022 est remarquable, et les processus gouvernementaux traditionnels ne sont pas en mesure de réagir à temps pour guider efficacement les technologues et protéger les citoyens. Les progrès rapides signifient également que même lorsque des lois sont proposées en temps opportun, elles peuvent rapidement devenir obsolètes. Deuxièmement, la grande incertitude concernant les capacités futures de l'IA rend difficile l'évaluation de ce que devraient être les garanties appropriées. C'est particulièrement le cas pour les conséquences d'une IA plus intelligente que l'homme, étant donné la réponse massive qui serait nécessaire pour une perturbation d'une telle ampleur. Troisièmement, 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 les modèles, ce qui met en péril l'éthique et la sécurité. Enfin, l'IA est un phénomène mondial, et 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. Pour relever ces défis, les gouvernements disposent d'une série d'options (tableau 2). La plus importante est de garantir la capacité et la structure internes pour surveiller les évolutions dans le domaine, y répondre, et se préparer aux prochaines vagues d'impacts.

Enfin, considérons notre positionnement stratégique : qu'est-ce que le Canada est particulièrement bien placé pour accomplir ? Nous n'avons pas la puissance militaire ou économique des États-Unis ou de l'Union européenne, et nous n'accueillons qu'une petite part des entreprises de pointe en IA, mais nous disposons de [talents](#) exceptionnels, d'une société relativement stable et prospère, et d'une bonne réputation sur la scène internationale. Le monde a besoin de champions de la gouvernance de l'IA, et il est temps pour nous de tirer parti de nos atouts et de montrer l'exemple. En agissant de cette manière, nous protégerons non seulement les Canadiens, nous renforcerons également le leadership du Canada sur ce qui est en train de devenir l'enjeu déterminant du XXI^e siècle.

Vitesse de développement de l'IA	➔	• Renforcer les capacités du gouvernement • Rendre la législation adaptable • Prévoir les impacts futurs de l'IA
Incertitude sur les risques à venir	➔	• Investir dans la recherche pour clarifier • Préparer les différents scénarios
Course au développement	➔	• Réglementation et contrôle • Harmoniser avec les lois étrangères
Échelle mondiale	➔	• Promouvoir des traités et des normes mondiaux
Risque d'étouffer les avantages de l'IA	➔	• Éviter les approches uniformes • Offrir des incitations à l'IA bénéfique

Tableau 2: Principaux défis et réponses possibles

Un plan pour le Canada: Cinq actions à mener en 2024/25

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 2025**:

1 Établir un ministère de l'IA

- ◆ Assure une voix au Conseil et une sensibilisation à l'impact de l'IA sur les principaux portefeuilles
- ◆ Peut héberger les agences de réglementation de l'IA, suivre les développements de l'IA et informer les dirigeants, communiquer avec les provinces et coordonner l'action du gouvernement fédéral

Raisonnement: L'IA est devenue trop importante pour être gérée par les ministères existants. Elle a besoin d'une voix directe au sein du Conseil des ministres pour assurer sa priorisation et la prise de conscience de son impact sur tous les autres portefeuilles principaux. Un ministère de l'IA est le meilleur endroit pour héberger les agences de contrôle et évitera à l'ISDE les conflits de mandat avec la promotion du développement. Il assure une responsabilité claire et peut grandir pour garantir que le gouvernement dispose des capacités nécessaires.

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

- ◆ Soutenir les efforts mondiaux pour lutter contre les effets de course au développement de l'IA
- ◆ Accueillir des discussions sur la gouvernance mondiale de l'IA lors de la présidence canadienne du G7 en 2025

Raisonnement: En l'absence de coordination mondiale, la réglementation nationale ne fera que repousser le développement de l'IA vers d'autres juridictions. De plus, le développement de l'IA de pointe se fait principalement à l'étranger, ce qui signifie que la réduction du risque systémique nécessite d'influencer les actions d'autres pays. Pour faire avancer les négociations mondiales, il faudra des champions nationaux, et le fait que le Canada accueille le G7 en 2025 est une occasion unique de faire preuve de leadership et d'avoir un impact positif.

3 Améliorer et adopter la Loi sur l'IA et les données

- ◆ Fournir un cadre juridique pour répondre aux préoccupations en matière d'éthique et de sécurité
- ◆ Améliorer l'indépendance en transférant l'administration au ministère de l'IA

Raisonnement: L'industrie et le public ont besoin d'un cadre juridique qui clarifie ce qui constitue un développement et une utilisation acceptables de l'IA. Étant donné que les dommages actuels se font déjà sentir et que les calendriers des capacités à venir sont incertains, il n'y a pas le temps d'abandonner la LIAD et de la réintroduire plus tard. [Nos recommandations détaillées sont disponibles sur le site GSIA.ca.](#)

4 Augmenter les investissements du Budget 2024 et les orienter vers la sécurité et la gouvernance

- ◆ Consacrer plus de 500 millions de dollars à la recherche sur l'IA technique et sur la politique de gouvernance, coordonnée par le nouvel Institut pour la sécurité de l'IA (ISIA) canadien
- ◆ Affecter plus de 20 % du fonds d'accès au calcul sur la sécurité de l'IA de pointe

Raisonnement: Le monde ne dispose actuellement ni des solutions techniques ni des mécanismes politiques efficaces pour garantir une IA sécuritaire et éthique. Avec des centaines de milliards investis chaque année pour augmenter ses 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 recherche en matière de gouvernance et de sécurité rattrape rapidement leur retard. Les 50 millions de dollars alloués à l'ISIA doivent être considérablement augmentés et le fonds d'accès au calcul doit être orienté vers la recherche en matière de sécurité.

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 de l'IA

Raisonnement: Même si les risques sécuritaires 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 dès maintenant. 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.