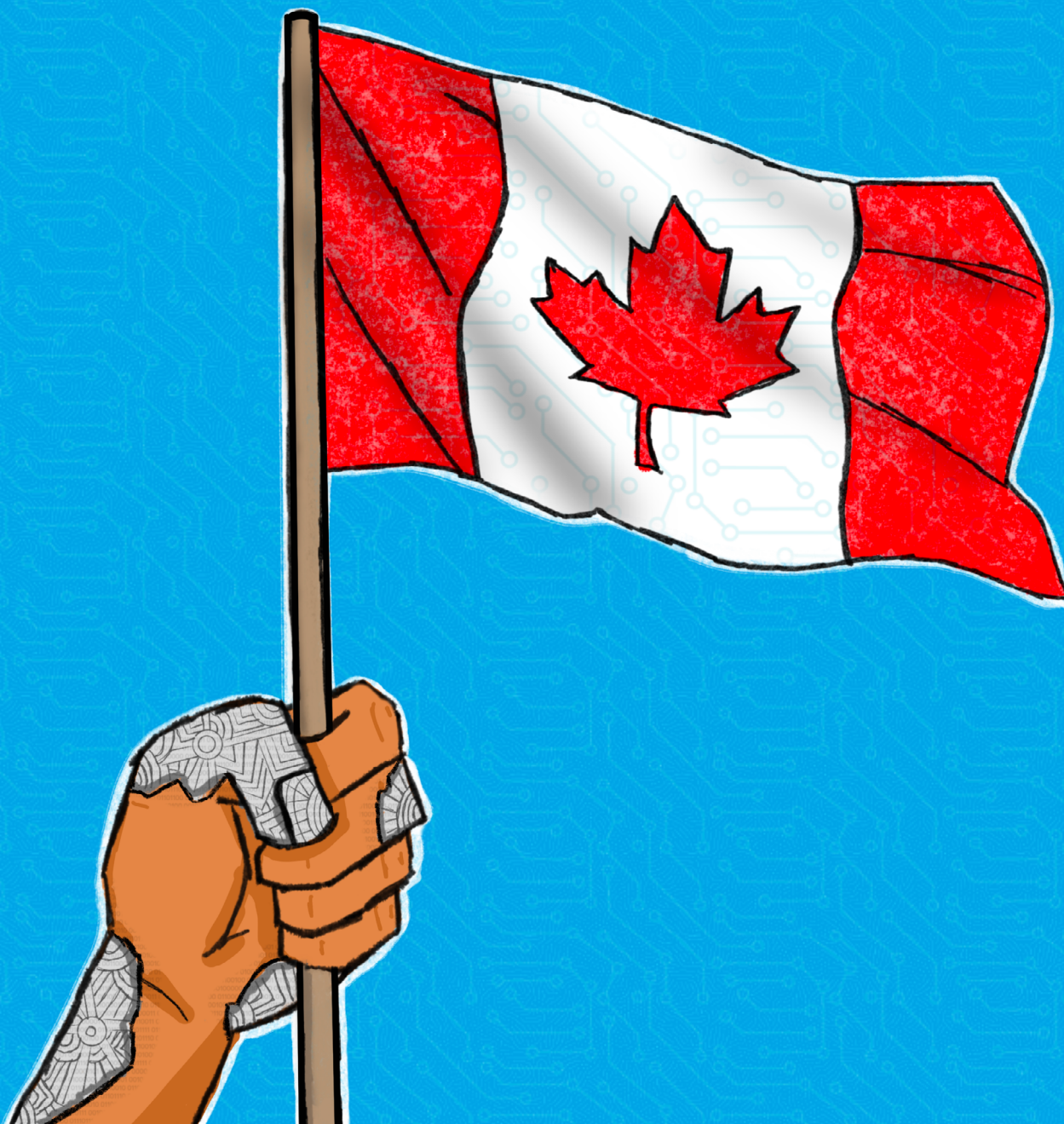


Qu'en est-il de l'informatique au Canada ?

Options politiques pour remédier au déficit informatique pour l'IA au Canada

Graham Dobbs et Jake Hirsch-Allen | Mars 2024



Remerciements



The Dais est une plateforme canadienne vouée à l'élaboration de politiques audacieuses et au développement de meilleures et meilleurs leaders. Nous sommes un institut d'idées sur les politiques publiques et de leadership à l'Université Toronto Metropolitan qui fait des liens entre les personnes, les idées et les pouvoirs dont nous avons besoin pour édifier un Canada plus inclusif, plus novateur et plus prospère.

Pour obtenir de plus amples renseignements, consultez le site dais.ca. 20, rue Dundas O., bureau 921, Toronto (Ontario) M5G 2C2

Pour citer le présent rapport :

Graham Dobbs et Jake Hirsch-Allen, « Qu'en est-il de l'informatique au Canada? Options politiques pour remédier au déficit informatique pour l'IA au Canada », the Dais, Mars 2024. <https://dais.ca>

© 2024, Université Toronto Metropolitan, 350, rue Victoria, Toronto (Ontario) M5B 2K3

L'institut The Dais est fier de mobiliser un groupe diversifié de bailleurs de fonds pour soutenir et catalyser notre travail, conformément à nos **valeurs**, et faisant l'objet d'un examen interne approfondi. En tant qu'institut non partisan d'intérêt public, nous acceptons uniquement les fonds provenant d'organismes qui soutiennent notre mission et nous permettent d'entreprendre des travaux de façon indépendante qui nous permet aussi d'exercer un contrôle éditorial total. Les noms de tous nos donateurs financiers sont affichés publiquement et de façon transparente sur tous les documents en ligne et imprimés pour chaque projet ou initiative.

Conception et illustration :

Mariana Rodrigues

Traduction :

Marie-Pierre Lavoie

Travail éditorial :

Suzanne Bowness, CodeWord Communications

Collaboratrices et collaborateurs :

Nina Rafeek Dow

André Côté

Viet Vu

Karim Bardeesy



Cet ouvrage est distribué sous licence en vertu d'une licence Creative Commons 4.0 – Attribution, pas d'utilisation commerciale, partage dans les mêmes conditions. Vous pouvez partager, copier ou redistribuer ce matériel, à condition : d'attribuer le crédit approprié; de ne pas l'utiliser à des fins commerciales; de ne pas appliquer de conditions légales ou de mesures technologiques qui empêchent légalement d'autres personnes de faire quelque chose qu'autorise cette licence; et si vous mélangez, arrangez ou adaptez le contenu, vous devez diffuser vos contributions sous les mêmes conditions que cette licence, indiquer si des modifications ont été apportées et ne pas suggérer que le concédant de la licence vous soutient ou soutient la façon dont vous avez utilisé son œuvre.



Auteurs

Graham Dobbs

Économiste principal

Comme économiste principal, Graham Dobbs (il) fait partie de l'équipe de recherche du Dais à TMU. Il explore les innovations éducatives et technologiques au sein de la population active canadienne et leur impact sur la répartition et les transitions professionnelles.

Par le passé, Graham a élaboré des outils et des idées destinés à rendre l'information sur le marché du travail plus accessible et navigable pour les Canadiennes et les Canadiens. Ses recherches antérieures ont porté sur le suivi des revenus imposables des personnes diplômées de l'enseignement supérieur et des compagnons du Sceau rouge afin de fournir de l'information crédibles sur les salaires aux étudiants, sur la cartographie de l'accès des adultes du Canada aux services d'orientation professionnelle et sur l'élargissement de l'accès aux compétences et à la demande d'emploi par le biais de tableaux de bord d'offres d'emploi en ligne.



Jake Hirsch-Allen

Conseiller principal

Jake Hirsch-Allen est conseiller principal à the Dais, où il se concentre sur les technologies et les compétences inclusives. Il établit des partenariats public-privé dans le domaine du développement de la main-d'œuvre et de l'enseignement supérieur pour LinkedIn. Les opinions exprimées dans ce document sont les siennes et celles de the Dais, mais pas celles de LinkedIn ou de Microsoft.

M. Hirsch-Allen conseille des investisseurs d'impact, des dirigeants du secteur public et de jeunes entreprises, notamment sur la technologie responsable et l'IA éthique. Il est membre du conseil d'administration d'Ontario Tech Talent et du Canadian Club et a fondé Lighthouse Labs. Ancien avocat spécialisé dans la propriété intellectuelle et le droit pénal international, M. Hirsch-Allen a également été président du comité technologique de la Global Education Platform, a enseigné la santé mondiale à l'Université McMaster et a été stagiaire à la Cour suprême d'Israël.



The Dais est un groupe de réflexion sur les politiques et le leadership de l'Université Toronto Metropolitan, qui se consacre à l'adoption d'une approche « technologie en premier » pour relever certains des défis les plus difficiles en matière d'économie, d'éducation et de démocratie, afin de promouvoir une prospérité et une citoyenneté partagées pour le Canada. Le présent rapport, qui s'appuie sur l'expertise de the Dais en matière de politique d'IA, a été lancé à la mi-décembre 2023, à la suggestion de notre conseil consultatif et après de nombreuses conversations avec des leaders de l'industrie et de la politique sur la nécessité d'une réaction urgente en matière de politique publique nationale à la question de la rareté de la capacité de calcul de l'IA au Canada.



Table des Matières

5 RÉSUMÉ

8 L'IA AU CANADA ET L'IMPORTANCE DE L'INFORMATIQUE

11 Qu'est-ce que l'intelligence artificielle et l'informatique pour l'intelligence artificielle ?

12 Financement public pour l'informatique pour l'IA au Canada

13 ÉTAT ACTUEL DE L'INFORMATIQUE POUR L'IA DANS LE MONDE ET AU CANADA

14 Capacité d'IIA au Canada :

17 Chaîne d'approvisionnement nationale en IIA

17 Efficacité de l'IIA au Canada

17 Efficacité de la recherche

18 Efficacité de l'industrie

19 Résilience de l'IIA au Canada

19 Politiques nationales pour l'expansion et l'appui de l'écosystème de l'informatique pour l'IA

19 Subventions et investissements internationaux pour les infrastructures pour l'IIA

22 Infrastructure de l'IIA comme moyen d'améliorer la gouvernance de l'IA

23 PRINCIPALES LEÇONS ET SCÉNARIOS POLITIQUES

23 Objectifs des politiques publiques et compromis

24 Scénario 1 : Centraliser et subventionner l'IIA par le biais de marchés publics fédéraux de la part de fournisseurs d'informatique en nuage

24 Scénario 2 : Travailler avec des partenaires commerciaux clés pour acheter conjointement l'IIA à grande échelle

26 Scénario 3 : Créer une capacité nationale de superinformatique pour l'IA

27 Considérations politiques pour tous les scénarios

30 CONCLUSION

1

Résumé

Le Canada est bien placé pour jouer un rôle de premier plan dans l'avenir de l'intelligence artificielle (IA). En 2017, le Canada a été le premier pays à lancer une stratégie nationale en matière d'IA, s'imposant ainsi comme un lieu d'accueil pour les chercheurs universitaires pionniers. Parmi les autres caractéristiques remarquables, citons un solide écosystème de jeunes entreprises dans le domaine de l'IA et l'une des communautés d'IA les plus qualifiées et à la croissance la plus rapide au monde. En l'adoptant plus largement, le pays pourrait tirer parti de l'IA pour remédier à ses problèmes de productivité de longue date.¹

Cependant, on s'inquiète de plus en plus du fait que, bien qu'il possède les talents et les algorithmes, le Canada est confronté à une pénurie du troisième intrant essentiel à la réalisation des ambitions nationales en matière d'IA : L'infrastructure informatique pour l'IA (IIA) (l'infrastructure physique comprenant des serveurs et des unités de traitement informatique pour former et développer des systèmes d'IA). Par rapport à ses pairs du G7, le Canada détient la plus faible quantité d'infrastructures et de performances informatiques

accessibles au public. Le potentiel actuel de performance informatique du Canada est inférieur de moitié à celui du Royaume-Uni, le pays du G7 le moins performant. Dans un environnement mondial hautement concurrentiel, ce déficit d'informatique pour l'IA pourrait empêcher les innovateurs canadiens de développer des entreprises d'IA prospères, ralentir l'adoption des technologies d'IA par les entreprises et limiter les progrès des chercheurs sur la prochaine génération d'enquêtes sur l'IA. Ce déficit constitue un risque pour la souveraineté du Canada en matière d'économie et de données dans un contexte mondial en pleine évolution.

Cependant, les ressources nécessaires sont considérables, voire immenses, et les avantages pour l'économie au sens large, sans parler de la société, restent hypothétiques. Pour citer une critique récente de la stratégie d'investissement d'une IIA publique américaine : « Nous devons savoir que ces investissements profiteront de manière significative à la société dans son ensemble, en élargissant l'horizon de l'innovation de manière à ce qu'elle profite au plus grand nombre de personnes et pas seulement à quelques-unes. »²

Par rapport à ses pairs du G7, le Canada détient la plus faible quantité d'infrastructures et de performances informatiques accessibles au public.

Le présent rapport sur les options politiques est le fruit d'un examen des données de recherche secondaires et de discussions approfondies avec des experts de la capacité informatique et de l'écosystème pour l'IA au Canada et il :

- **définit le contexte de l'IA au Canada et explique pourquoi la capacité informatique pour l'IA est importante.**
- **résume l'état de l'informatique pour l'IA au Canada** par rapport à des pays homologues, comme on l'analyse grâce à la lentille du cadre national IIA de l'OCDE.
- **présente trois scénarios et des approches fondées sur les politiques publiques,** y compris les compromis qui en découlent, pour remédier au déficit informatique pour l'IA au Canada.

Le présent rapport fournit de l'information clé sur l'état actuel de l'IIA au Canada, dans trois domaines principaux :

- **Capacité d'IIA :** Le Canada se trouve derrière tous les autres pays du G7 en matière d'infrastructures d'IIA. L'infrastructure d'IIA dont il dispose est insuffisante pour la recherche exploratoire.
- **Efficacité de l'IIA :** Le Canada utilise principalement sa capacité d'IIA pour la recherche, tandis que les entreprises privées doivent s'appuyer sur des solutions basées sur l'informatique dématérialisée (grandes multinationales technologiques comme Amazon et Microsoft) à l'extérieur du Canada. Les pressions liées au prix limitent leur utilisation.
- **Résilience de l'IIA :** Contrairement à de nombreux pays homologues, la politique publique canadienne repose sur de vastes programmes qui encouragent l'adoption de la technologie en général, sans politiques ciblant spécifiquement l'accès à l'IIA.

Nous supposons que le gouvernement du Canada et les autres intervenants de l'écosystème de l'IA voudront adopter cet objectif de politique publique pour accroître l'accès à l'IIA, remédier au déficit informatique pour l'IA et, par conséquent, aider à retenir les entreprises et les talents d'IA et promouvoir un plus grand développement et une plus grande adoption de l'IA. À son tour, cet objectif peut promouvoir un objectif plus large de politique publique visant à accroître la productivité et la concurrence nationales. Comme Internet avant elle, la capacité d'IA pourrait alimenter un large éventail de domaines de la croissance économique qui nécessitent des solutions immédiates, à moyen et à long terme. L'IIA a le potentiel d'être traitée comme une infrastructure de type utilitaire.

Dans ce contexte, nous présentons trois scénarios politiques comme autant d'approches possibles :

- **Scénario 1 : Centraliser et subventionner l'IIA par le biais de marchés publics fédéraux ou de subventions directes de la part des fournisseurs d'informatique en nuage existants**
- **Scénario 2 : Travailler avec des partenaires commerciaux clés pour acheter conjointement l'infrastructure IIA et l'informatique en nuage à grande échelle et y accéder**
- **Scénario 3 : Créer une capacité nationale de superinformatique pour l'IA par l'intermédiaire de fournisseurs de premier plan ou de fournisseurs tiers en vertu de partenariats pluriannuels et multifournisseurs**

Le réseau d'acteurs canadiens ayant besoin d'un accès à l'informatique pour l'IA comprend des chercheurs, des organismes sans but lucratif, des gouvernements, des entreprises axées sur l'IA qui forment des modèles d'IA et des entreprises qui ont besoin d'un accès à la puissance de l'informatique pour exécuter des applications d'IA. Si ces acteurs ont besoin de l'informatique d'IA, nous pouvons construire cette capacité au niveau national ou permettre l'accès à la capacité internationale, car l'informatique pour l'IA au Canada ne sera tout simplement pas disponible en quantités suffisantes selon l'échelle de temps selon laquelle les utilisateurs ont besoin d'y accéder. Nous avons donc besoin de stratégies à court (six prochains mois) et à moyen terme (deux prochaines années) pour ouvrir l'accès, tout en nous efforçant d'augmenter la capacité d'informatique pour l'IA au Canada.

Même à long terme, la souveraineté en matière d'informatique pour l'IA n'est peut-être pas totalement réalisable, et le jeu n'en vaut peut-être pas la chandelle : nous continuerons à être interdépendants des chaînes d'approvisionnement et des fournisseurs hébergés à l'étranger. Si l'accès au marché peut être garanti, les acteurs du secteur privé en particulier doivent s'attendre à continuer à chercher à l'extérieur comme à l'intérieur du Canada un accès à l'informatique pour l'IA.

La technologie nécessaire à l'informatique pour l'IA et les besoins des futurs modèles et applications de l'IA pourraient être très différents de ce que nous connaissons aujourd'hui. Dans tous les cas possibles, le Canada doit agir maintenant pour regagner une partie de sa position, tout en restant agile et prêt à s'adapter à l'évolution de la situation.

Une idée audacieuse :

Le Canada doit s'attaquer d'urgence au déficit informatique pour l'IA afin de conserver et d'accroître ses investissements dans les entreprises et les talents, et améliorer sa productivité.



2


L'IA au Canada et l'importance de l'informatique

Le Canada est reconnu internationalement pour sa contribution à l'intelligence artificielle (IA), car il a été l'un des premiers pays à adopter une législation et à établir une politique nationale et un cadre réglementaire pour l'IA, des efforts étant actuellement déployés pour élaborer des normes industrielles harmonisées à ce qui se fait à l'international pour l'IA. Le Canada a également été le premier pays à lancer un plan national pour le développement de l'IA, la Stratégie pancanadienne en matière d'IA (SPCIA), dirigée par l'Institut canadien de recherches avancées (CIFAR). Ces initiatives, parmi d'autres, ont contribué à faire du Canada un leader en matière de recherche, de talents et de croissance des jeunes pousses dans le domaine de l'IA.

Le pays compte des pionniers dans le domaine de l'IA, notamment Geoffrey Hinton et Yoshua Bengio, qui ont beaucoup travaillé sur les algorithmes de réseaux neuronaux³ et Richard Sutton dans le domaine de l'apprentissage par renforcement. Ces noms figurent parmi les plus en vue d'un solide écosystème de chercheurs, qui est en tête des pays du G7 en termes

de publications de recherche sur l'IA par habitant en date de mai 2023.⁴ La main-d'œuvre canadienne en IA est estimée entre 200 000 et 220 000 personnes et se compose de professionnels de la science des données, du génie logiciel et du développement; cela représente un pour cent de la main-d'œuvre canadienne en 2023 (contre seulement 0,5 % de la main-d'œuvre aux États-Unis en 2019).^{5,6} D'autres études suggèrent également la position de leader du Canada en matière de talents dans le domaine de l'IA.^{7,8}

Le Canada peut également se vanter de compter un écosystème dynamique de jeunes entreprises en IA, qui comprend plus de 1 500 entreprises d'IA, et plus de 150 entreprises ont récolté un total de 2,5 milliards de dollars en capital-risque en 2023.^{9,10} Le Canada se classe au quatrième rang des dix premières nations en matière d'investissement en capital-risque, de financement total mobilisé, de création d'entreprises d'IA générative (IAG) et de croissance du nombre de dépôts de brevets d'IA par habitant en date de la mi-2023.¹¹



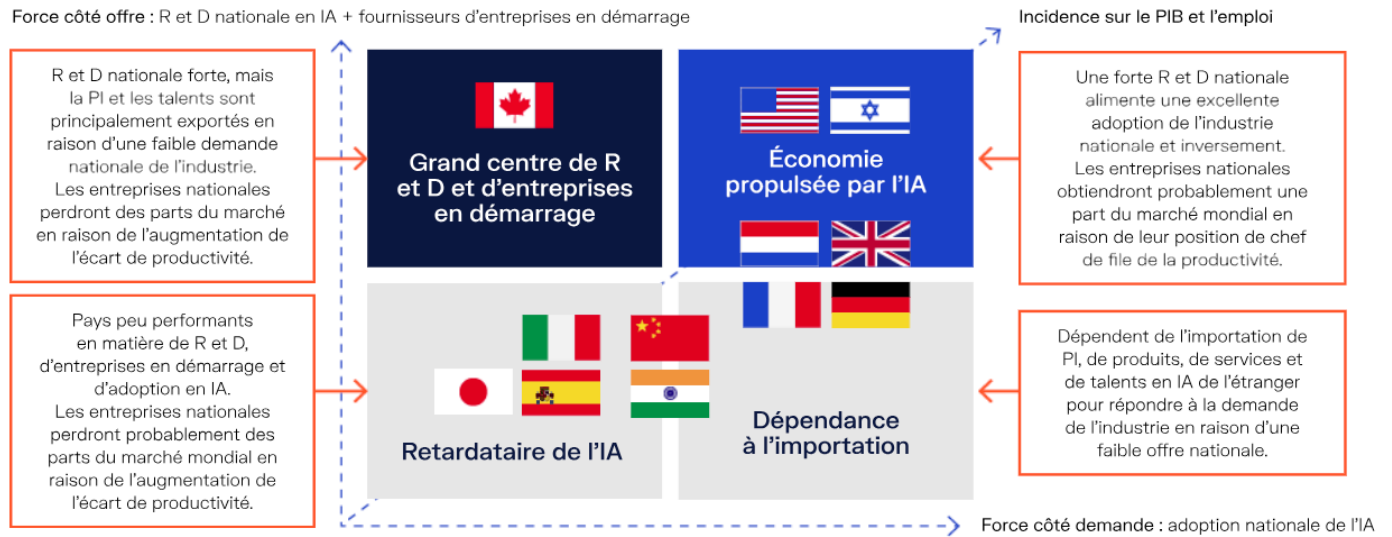
Tandis que les chercheurs et les entreprises du Canada sont confrontés à une course au développement et à l'adoption de l'IA dans un domaine qui progresse rapidement – et que les décideurs politiques s'efforcent de trouver le juste équilibre entre les conditions favorables et les garde-fous – une question essentielle a reçu relativement peu d'attention au pays : la capacité « informatique pour l'IA » (IIA) du Canada.

La stratégie canadienne en matière d'IA est financée par le ministère Innovation, Sciences et Développement économique Canada (ISDE) du gouvernement fédéral en vertu de la Stratégie pancanadienne en matière d'intelligence artificielle (SPCIA, ou la Stratégie). Établie en 2017, la stratégie a alloué plus de 573 millions de dollars pour appuyer les organismes d'IA afin de se concentrer sur la commercialisation, les normes, les talents et la recherche jusqu'en 2031. Le CIFAR, appuyé par ISDE, administre la SPCIA avec 160 millions de dollars depuis 2017 et a reçu une somme de 208 millions de dollars supplémentaire en 2022. Le CIFAR appuie les instituts de recherche en IA comme l'Alberta Machine Intelligence Institute (AMII), le Montreal Institute for Learning Algorithms (Mila) et le Vector Institute ainsi que les pôles d'innovation mondiaux du Canada, notamment Scale AI, et le Conseil canadien des normes dans l'élaboration de normes d'adoption de l'IA.

Tandis que les chercheurs et les entreprises du Canada sont confrontés à une course au développement et à l'adoption de l'IA dans un domaine qui progresse rapidement – et que les décideurs politiques s'efforcent de trouver le juste équilibre entre les conditions favorables et les garde-fous – une question essentielle a reçu relativement peu d'attention au pays : la capacité « informatique pour l'IA » (IIA) du Canada.

Les jeunes entreprises canadiennes spécialisées dans l'IA sont confrontées à des difficultés pour passer de l'innovation à la commercialisation¹² Des recherches antérieures ont montré qu'en dépit d'un solide écosystème de jeunes entreprises, l'adoption de l'IA par les entreprises générales est faible : moins d'une entreprise canadienne sur 25 déclare utiliser l'IA dans ses activités, ce qui est l'un des taux les plus bas des 35 pays de l'OCDE et le deuxième taux le plus bas des pays du G7.¹³ Un rapport récent de Scale AI, le pôle mondial d'innovation en IA du Canada, constate que le Canada se trouve à un moment critique : malgré la vigueur de la recherche et de la RetD nationales et des jeunes entreprises d'IA, le Canada manque d'ingrédients clés pour permettre à l'IA d'offrir de manière responsable des avantages à l'économie. Le Canada risque de se laisser distancer par ses pairs internationaux qui traduisent la RetD et l'adoption de l'IA par l'industrie en améliorations de la productivité, ce qui affaiblirait la concurrence du pays.

Figure 1. Graphique du rapport Portrait de l'IA d'ici qui décrit la trajectoire de l'écosystème de l'IA au Canada¹⁴



Des rapports récents suggèrent que les ressources d'IIA représentent une part importante des coûts d'investissement d'une entreprise d'IA, jusqu'à la moitié des dépenses d'investissement annuelles de nombreuses entreprises.^{15 16} Même les plus grandes entreprises, bien qu'elles dépensent des dizaines de millions en infrastructure informatique, n'ont pas les moyens de s'offrir l'infrastructure informatique spécifique à l'IA la plus rapide et à l'informatique en nuage, et encore moins d'y accéder.¹⁷ En février 2023, l'OCDE a souligné que l'absence de plans ciblés pour la capacité nationale d'informatique pour l'IA dans la stratégie nationale en matière d'IA constituait un « angle mort [qui] pourrait mettre en péril les objectifs économiques nationaux ». ¹⁸ Tandis que les décideurs politiques à tous les niveaux ont commencé à se pencher sur la question, l'infrastructure nationale de superinformatique du Canada est à la traîne en ce qui concerne le nombre d'unités informatiques disponibles (capacité) et les performances (efficacité) en matière de formation et d'inférence de l'IA.

Qu'est-ce que l'intelligence artificielle et l'informatique pour l'intelligence artificielle ?

Avant d'aller plus loin, examinons la définition de l'intelligence artificielle (IA) et de l'informatique pour l'intelligence artificielle (IIA) :

Qu'est-ce que l'intelligence artificielle (IA) ?

Selon l'Organisation de coopération et de développement économiques (OCDE)¹⁹, l'intelligence artificielle (IA) « est un système qui fonctionne grâce à une machine et capable d'influencer son environnement en produisant des résultats (tels que des prédictions, des recommandations ou des décisions) qui peuvent influencer les environnements physique réels ou virtuels. Les systèmes d'IA sont conçus pour fonctionner de façon plus ou moins autonome et adaptable après leur déploiement. »²⁰

Les systèmes d'IA ont besoin de puissance informatique pour générer des inférences à partir de leurs données d'entrée, qui sont généralement effectuées par l'un de deux types de systèmes :

(a) les infrastructures informatiques parallèles comme les superordinateurs, y compris les superordinateurs publics utilisés principalement à des fins de recherche universitaire et scientifique ou

(b) les systèmes informatiques en nuage, généralement exploités par de grandes sociétés privées américaines cotées en bourse, comme Microsoft (Azure), Google (Cloud) et Amazon (Web Services). Pour combiner ces multiples contextes, nous nous appuyons sur la définition du groupe de travail de l'OCDE²¹ sur l'informatique pour l'intelligence artificielle (IIA) :

Qu'est-ce que l'informatique pour l'IA (IIA) ?

L'informatique pour l'IA est une pile spécialisée de matériel et de logiciels qui comprend des processeurs ou des puces, des serveurs, du stockage, des logiciels et des réseaux, tous conçus pour prendre en charge des charges de travail et des applications spécifiques à l'IA. L'informatique pour l'IA couvre un éventail de technologies différentes, des puces informatiques aux serveurs de données en passant par l'informatique en nuage.

Le développement de l'IA comporte deux phases : la formation et l'inférence. Au cours de la première phase, un modèle d'IA est « formé » à partir de données. Au cours de la deuxième phase, un modèle d'IA formé est déployé et « infère » (c'est-à-dire qu'il prend des décisions ou des mesures) sur le terrain en fonction des nouvelles données. Les phases de formation et d'inférence sont exécutées sur une infrastructure informatique pour l'IA dans un centre de données privé ou dans une infrastructure en nuage accessible au public (comme Amazon Web Services).


Nous développerons cette définition dans les sections suivantes.

Financement public pour l'informatique pour l'IA au Canada

Sur les 573 millions de dollars de financement de la Stratégie pancanadienne en matière d'IA, seuls 40 millions sont directement affectés aux initiatives d'informatique pour l'IA. La majorité de l'infrastructure informatique pour l'IA relève de la responsabilité de l'Alliance de recherche numérique du Canada (ARNC) et de ses partenaires de recherche. L'ARNC est le successeur de l'écosystème informatique de haute performance qui a vu le jour dans les années 1990 et qui a évolué grâce à des initiatives axées sur la recherche comme SHARCNET et Calcul Canada. L'ARNC est désormais financée par la SPCIA et collabore avec les opérateurs du Réseau national de la recherche et de l'éducation, CANARIE (autrefois le Réseau canadien pour l'avancement de la recherche, de l'industrie et de l'enseignement), afin de fournir des ressources informatiques aux chercheurs, aux jeunes entreprises et aux établissements d'enseignement.

L'investissement dans l'infrastructure informatique du Canada a été largement limité au gouvernement et aux institutions de recherche. En outre, les performances informatiques médiocres de l'infrastructure actuelle empêchent effectivement de traiter les charges de travail liées à l'IA à l'intérieur des frontières nationales.

Il n'existe aucun programme canadien axé sur les besoins en informatique pour l'IA des acteurs du secteur privé, en particulier des entreprises qui cherchent à faire évoluer ou à développer considérablement l'infrastructure publique d'informatique pour l'IA. Le Canada est déjà en retard, comme nous le verrons dans la section suivante, et il s'agit donc d'un déficit politique auquel il faut remédier de toute urgence.



Il n'existe aucun programme canadien axé sur les besoins en informatique pour l'IA des acteurs du secteur privé, en particulier des entreprises qui cherchent à faire évoluer ou à développer considérablement l'infrastructure publique d'informatique pour l'IA.

3

État actuel de l'informatique pour l'IA dans le monde et au Canada

Les entreprises et les chercheurs canadiens dans le domaine de l'IA sont confrontés à un déficit informatique pour l'IA, et les mesures de politique publique pour y remédier sont insuffisantes.

Le manque d'accès et la pénurie de matériel spécifique à l'IIA, notamment les unités centrales graphiques et les unités de traitement de tenseur (GPU/TPU²²), limitent le potentiel de l'écosystème d'IA du Canada à prospérer dans un paysage mondial hautement concurrentiel.

Bien que les besoins spécifiques en informatique pour l'IA puissent changer à l'avenir,²³ un nombre croissant de chefs d'entreprise, d'experts en innovation et de

chercheurs en IA ont soulevé le manque d'accès à l'IIA au Canada comme une contrainte commerciale majeure.

Pour évaluer l'état de l'IIA au Canada, nous examinons le cadre stratégique de l'OCDE pour le renforcement des capacités informatiques nationales, qui développe la définition de l'OCDE présentée dans la section précédente.²⁴ Le cadre détermine trois dimensions principales de l'IIA nationale :

Dimensions nationales de l'IIA de l'OCDE

Capacité

- Quelle est la disponibilité et l'utilisation de l'IIA nationale ?
- Combien d'IIA nationale utilise-t-on, qui l'utilise et dans quels secteurs l'utilise-t-on ?

Efficacité

- Quelle est l'efficacité de l'utilisation des capacités nationales d'IIA ?
- Existe-t-il suffisamment de main-d'œuvre qualifiée, de RetD, d'accès abordable et un environnement politique favorable ?

Résilience

- Quelle est la résilience de la capacité informatique d'un pays (p. ex., sécurisée, souveraine, durable) ?
- Qui possède la capacité et où se trouve-t-elle ? Les chaînes d'approvisionnement sont-elles sûres ?



Bien que les besoins spécifiques en informatique pour l'IA puissent changer à l'avenir, un nombre croissant de chefs d'entreprise, d'experts en innovation et de chercheurs en IA ont soulevé le manque d'accès à l'IIA au Canada comme une contrainte commerciale majeure.

Le cadre stratégique de l'OCDE fournit une grille d'analyse qui permet d'identifier les principales lacunes, forces et occasions du futur écosystème d'informatique pour l'IA du Canada. Notre analyse utilise l'approche suivante pour appliquer les dimensions du cadre de l'OCDE :

- La **capacité de l'IIA** est l'infrastructure informatique mise à la disposition de la recherche et de l'industrie, mesurée par le nombre de noyaux informatiques et la puissance informatique disponibles.
- L'**efficacité de l'IIA** examine qui, où et comment l'infrastructure informatique pour l'IA nationale est utilisée à des fins industrielles et de recherche.
- La **résilience de l'IIA** est la stratégie politique actuelle, les partenariats internationaux et/ou industriels, et garantit un accès cohérent et fiable aux ressources informatiques dans des scénarios perturbateurs (y compris les catastrophes économiques, géopolitiques et naturelles).

Nous explorons chacun de ces trois aspects au Canada.

Capacité d'IIA au Canada :

La mesure de base pour l'évaluation et l'étalonnage de la capacité de calcul nationale (infrastructure pour l'IIA et supercalculateurs) est la performance des opérations en virgule flottante par seconde (FLOPS).²⁵ En novembre 2023, la capacité du Canada s'élèvera à 41 pétaFLOPS (PFLOPS, ou 1 quadrillion FLOPS), soit 0,7 % de la performance mondiale en matière de informatique. À titre de comparaison, les États-Unis disposent de 3 700 PFLOPS, soit 53 % de la capacité informatique mondiale, le Japon arrive en deuxième position avec 670 PFLOPS, soit 10 %, suivi de la Chine avec 407 PFLOPS, soit 6 %.

Pour mieux cerner les déficits de capacité et de performance informatique entre les pays, nous calculons le ratio par habitant et ajusté au PIB du Canada par rapport à ses homologues du G7. Nous multiplions ensuite les ratios d'ajustement par les paramètres IIA du Canada afin de comparer plus efficacement les paramètres de performance totale de la capacité informatique des comparateurs internationaux. Le déficit en matière d'informatique de l'IA persiste même si l'on tient compte des différences entre les marchés internationaux et les populations. En termes de comparaison, les États-Unis devancent le Canada en termes de capacités IIA ajustées d'un facteur de 8 à 11 fois, le Japon d'environ 8 fois et la France et l'Allemagne de 2 à 3 fois. Concrètement, cela signifie que l'écosystème canadien de l'IA a moins accès à la puissance informatique au niveau national que ses homologues internationaux.

Compte tenu de la force relative du Canada par rapport aux pays de l'OCDE et du G7 en ce qui concerne d'autres paramètres liés à l'IA, ce déficit informatique pour l'IA crée des défis particuliers pour les entreprises nationales d'IA à forte croissance du Canada et les travailleurs qui possèdent des compétences en matière d'IA. Il ne s'agit pas seulement d'un écart de concurrence entre le Canada et ses homologues, mais aussi d'un écart à l'intérieur du pays, où les talents, les entreprises et la recherche à forte capacité se heurtent à

une infrastructure informatique à faible capacité. Les entreprises qui ont besoin d'IIA achètent ces ressources sur le marché libre auprès de fournisseurs de services en nuage principalement basés aux États-Unis. Les entreprises et les chercheurs canadiens spécialisés dans l'IA sont contraints de se tourner vers d'autres pays pour bénéficier d'incitations à l'innovation et à la commercialisation, ce qui laisse le Canada encore plus en retard sur le marché très concurrentiel de la productivité de l'IA et de son potentiel de croissance économique.

Compte tenu de la force relative du Canada par rapport aux pays de l'OCDE et du G7 en ce qui concerne d'autres paramètres liés à l'IA, ce déficit informatique pour l'IA crée des défis particuliers pour les entreprises nationales d'IA à forte croissance du Canada et les travailleurs qui possèdent des compétences en matière d'IA.

Tableau 1. Capacité informatique du Canada par rapport à ses pairs²⁶

Pays	Capacité de performance informatique non ajustée	Performance par habitant	Performance ajustée selon le PIB	Performance de base (PFLOPS)	Ratio de population	Ratio du PIB
États-Unis	90,4	10,6	7,6	3 725,85	8,56	11,90
Japon	16,3	7,6	8,5	669,83	2,15	1,91
Italie	8,5	5,6	8,9	351,76	1,51	0,96
Allemagne	6,2	1,9	3,1	256,27	3,21	1,98
France	4,2	2,4	3,2	173,23	1,75	1,30
Royaume-Uni	2,0	1,2	1,4	81,71	1,72	1,44
Canada	1,0	1,0	1,0	41,21	1,00	1,00

Notamment, cette contrainte est encore plus forte pour les meilleurs talents qui travaillent sur des modèles d'avant-garde. Seuls trois des cinq superordinateurs universitaires canadiens disponibles disposent du matériel spécifique requis pour une informatique efficace pour l'IA. Avec des capacités, des performances et du matériel d'IA spécialisé

(GPU) limités, les écosystèmes d'IA canadiens sont confrontés à des temps de formation plus longs et à des mesures d'allocation des ressources plus strictes lorsqu'ils explorent des innovations d'IA d'avant-garde, par rapport à d'autres pays leaders en matière d'innovation dans le domaine de l'IA.²⁷

L'infrastructure publique pour l'IA du Canada doit améliorer ses performances d'un ordre de grandeur pour satisfaire aux besoins des innovations d'avant-garde en matière d'IA, comme le transformateur génératif pré-formé version 3 (GPT-3) d'OpenAI. À titre d'exemple, GPT-3 nécessite 3 640 PFLOPS de performance informatique pour se former en une seule journée.²⁸ Pour mettre les choses en perspective, la combinaison de toutes les ressources informatiques canadiennes permettrait de former ce modèle en 100 jours. Si l'on ne peut utiliser que les cinq superordinateurs universitaires, la formation prendrait six mois. En comparaison, il faudrait moins de quatre jours au supercalculateur de recherche américain Frontier et environ six jours et demi au supercalculateur privé Azure Eagle de Microsoft pour effectuer la même charge de travail.

Les infrastructures qui appartiennent à l'industrie restent également rares pour le déploiement commercial de l'IA au Canada. Des signes indiquent que ce phénomène est en train de se développer, même si les bénéfices reviendront principalement aux entreprises basées aux États-Unis.²⁹ Amazon Web Services prévoit d'investir plus de 24,8 milliards de dollars dans l'infrastructure en nuage au Canada jusqu'en 2037, et a déjà investi 2,57 milliards de dollars dans le pays. Ce montant est à comparer aux 40 millions de dollars de contribution à l'infrastructure pour l'IA réservés à l'ARNC en vertu du financement de la SPCIA 2022.

Tableau 2. Supercalculateurs nationaux et emplacements³⁰

Supercalculateurs et hôtes nationaux	Noyaux	Rmax (PFlop/s)	Rpeak (PFlop/s)	Puissance (kW)
Underhill – Services partagés Canada	148 320	7,76	10,92	1 295
Robert – Services partagés Canada	148 320	7,76	10,92	1 295
Narval – Calcul Québec / Compute Canada	76 320	5,89	12,17	311
Niagara – SciNet / Université de Toronto / Compute Canada	80 640	3,6	6,25	919
Cedar (GPU) – Université Simon-Fraser / Compute Canada	67 584	3,37	5,83	310
Banting – Services partagés Canada	53 200	2,68	4,09	
A14A – Software Company MCA	78 336	2,66	5,26	
Cedar (CPU) – Université Simon-Fraser / Compute	67 584	2,61	4,9	792
Daley – Services partagés Canada	53 200	2,6	4,09	
Béluga – Calcul Québec / Compute Canada	72 480	2,28	7,49	240
Total	845 984	41,21	71,92	5 162
Gouvernement	403 040	20,8	30,02	2 590
Recherche	364 608	17,75	36,64	2 572
Privé	78 336	2,66	5,26	S.O.



La chaîne d'approvisionnement nationale en matière d'IA, qui n'en est qu'à ses balbutiements, entrave également le renforcement des capacités en matière d'IIA au Canada.

Pourtant, l'investissement privé dans la construction d'IIA au Canada reste en retard par rapport aux États-Unis. Amazon prévoit de dépenser 47 milliards de dollars pour de nouveaux centres de données rien qu'en Virginie d'ici 2040.³¹ Tandis que Microsoft, Google et Amazon commencent à développer les capacités canadiennes en matière d'informatique dématérialisée, les entreprises canadiennes en phase de démarrage n'ont pas les moyens d'y accéder.

Chaîne d'approvisionnement nationale en IIA

La chaîne d'approvisionnement nationale en matière d'IA, qui n'en est qu'à ses balbutiements, entrave également le renforcement des capacités en matière d'IIA au Canada. Le Canada héberge plus de 100 entreprises qui mènent des activités de RetD dans le domaine des semi-conducteurs pour les véhicules et l'IA générative. Les principaux acteurs du secteur de l'IA, comme Radical Ventures, Ada et Cohere, illustrent la croissance substantielle du déploiement de capital-risque dans les entreprises canadiennes d'IA. Les entreprises canadiennes spécialisées dans les infrastructures d'IA et le matériel informatique, comme Ranovus, Tenstorrent, Untether, Tartan AI et Epic Semiconductors, fabriquent des infrastructures en nuage, des noyaux informatiques pour serveurs et du matériel d'inférence spécialisé dans l'IA.³² Récemment, Tenstorrent, un concepteur de puces d'IA, a concédé une licence sur sa propriété intellectuelle à Rapidus, une jeune entreprise japonaise de puces d'IA appuyée par l'État, qui a investi 2 milliards de dollars dans la fabrication de puces à Hokkaido.³³ Si ces investissements sont louables et reflètent les tendances observées dans les investissements privés dans l'IIA, les grandes économies concurrentes ont investi plus massivement dans les chaînes d'approvisionnement de l'IA, de la conception des puces à la distribution, en passant par l'infrastructure et les services commerciaux en nuage, pour ainsi offrir de plus grandes occasions de marché et une infrastructure d'IIA rentable.³⁴

À l'international, nous observons certains signes de consolidation des entreprises de la chaîne d'approvisionnement de l'informatique pour l'IA sous et dans les grands fournisseurs d'IIA nuagique. Le tableau 4 présente des exemples de cette consolidation au niveau national.

Efficacité de l'IIA au Canada

Efficacité de la recherche

L'importante infrastructure d'IIA au Canada est destinée aux chercheurs universitaires et gouvernementaux dans les domaines de la science biomédicale, de la science du climat et de l'ingénierie aérospatiale.³⁵ Le Canada héberge actuellement 10 sites informatiques à grande échelle, dont quatre sont déployés par Services partagés Canada (SPC, 21 PFLOPS ou 48 % de la capacité totale), un est exploité par la société de conseil en ingénierie MCA (9 %) et cinq sites sont répartis entre des universités régionales (18 PFLOPS ou 43 %).³⁶ La majorité de l'industrie canadienne de l'IA et des utilisateurs précoces accèdent à l'IIA par l'intermédiaire de fournisseurs de services d'informatique en nuage basés aux États-Unis.

Bien qu'il ait accès à près de la moitié de l'infrastructure nationale d'IIA, l'écosystème de recherche canadien est confronté à une pénurie de ressources informatiques par rapport aux écosystèmes internationaux. Les superordinateurs de recherche les plus puissants du Canada, Narval et Béluga, qui contiennent 76 000 cœurs de calcul, sont dérisoires par rapport aux 8,7 millions de cœurs de calcul du superordinateur américain Frontier, ou au million de cœurs du superordinateur Eagle de Microsoft.³⁷

Les superordinateurs de recherche les plus puissants du Canada, Narval et Béluga, qui contiennent 76 000 cœurs de calcul, sont dérisoires par rapport aux 8,7 millions de cœurs de calcul du superordinateur américain Frontier, ou au million de cœurs du superordinateur Eagle de Microsoft.

Efficacité de l'industrie

Étant donné que la majorité de la capacité nationale de superinformatique est destinée à la recherche, l'industrie canadienne de l'IA, les autres petites et moyennes entreprises (PME) et les grandes entreprises qui cherchent à adopter l'IA sont principalement tributaires de la location d'une capacité limitée informatique en nuage pour l'IA auprès de grands fournisseurs comme Amazon Web Services (AWS), Google Cloud Platform et Microsoft Azure à l'extérieur des frontières du Canada.³⁸

À titre de comparaison, l'entreprise américaine Compute détient plus d'un quart de la capacité totale informatique (27 % contre 9 % pour le Canada) et de performance (31 % contre 7 % pour le Canada).

Cela pose des problèmes d'évolutivité nationale et de souveraineté économique et des données.³⁹

L'absence d'infrastructure en nuage d'entreprise nationale spécifique à l'IA incite les entreprises en phase de démarrage à opérer en dehors des frontières canadiennes, ce qui pose des problèmes de souveraineté des données et de rétention des talents.

Les coûts élevés de l'informatique pour l'IA restent un obstacle pour les premières entreprises d'IA. Le fondateur d'OpenAI, Sam Altman, a récemment estimé à plus de 130 millions de dollars le coût de la formation du modèle d'IA à l'origine de GPT-4.⁴⁰ Bien qu'il existe d'autres formes de modèles d'IA frontaliers moins gourmands en ressources informatiques, les prix actuels de l'IIA restent intenable pour la majorité des entreprises nationales d'IA. Les entreprises nationales d'IA en phase de démarrage sont confrontées à des compromis dans l'échange de capitaux avec les opérateurs historiques de l'informatique en nuage afin d'acquérir les ressources nécessaires pour proposer des offres d'IA efficaces et commercialement viables.⁴¹ Les gagnants sont les géants de l'industrie de l'IA comme Microsoft, Google et Amazon, qui sont payés pour la croissance de presque tous les besoins informatiques.⁴²

L'absence d'IIA publique ne laisse aux écosystèmes et à l'industrie de l'IA nationale aucune autre option viable que l'externalisation de leurs activités commerciales et de leurs données vers l'infrastructure d'IIA de l'industrie privée américaine.⁴³

Tableau 3. Infrastructure en nuage d'un fournisseur de services basé aux États-Unis et investissements au Canada

Emplacement de l'infrastructure en nuage	Investissement depuis 2014	Toronto	Québec City	Montréal	Calgary
Microsoft ⁴⁴	environ 677 millions de \$ ⁴⁵	x	x		
Google ⁴⁶	environ 735 millions de \$ ⁴⁷	x		x	
Amazon	environ 2 570 millions de \$ ⁴⁸			x	x

Les coûts élevés de l'informatique pour l'IA restent un obstacle pour les premières entreprises d'IA.

Résilience de l'IIA au Canada

Politiques nationales pour l'expansion et l'appui de l'écosystème de l'informatique pour l'IA

En 2022, la Stratégie pancanadienne en matière d'IA (SPCIA) a alloué 40 millions de dollars à l'expansion de la capacité informatique nationale pour son écosystème de recherche.⁴⁹ Le gouvernement du Canada offre un certain nombre de programmes aux entreprises canadiennes pour qu'elles adoptent des capacités d'IA par le biais de divers programmes ainsi que de financements axés sur la recherche et l'innovation par l'intermédiaire du CIFAR. Il s'agit de subventions indirectes comme des incitations fiscales, des compensations salariales, des petites subventions et des prêts aux entreprises. Quelques exemples de programmes d'appui à l'industrie sont résumés ci-dessous :

- La recherche scientifique et le développement expérimental (RSetDE) offrent une incitation fiscale pour les dépenses liées à la RetD dans l'avancement de l'IA et des services de conseil sur l'optimisation du potentiel de revendication.
- Le Programme canadien d'adoption du numérique (PCAN) offre des subventions pour des services de conseil en matière d'adoption et de transformation numériques et des ressources en matière de cybersécurité.
- Les programmes d'apprentissage intégré au travail (AIT) numérique subventionnent les salaires des jeunes diplômés et des étudiants du postsecondaire pour les entreprises et les organismes spécialisés dans l'IA.
- Des prêts sans intérêt de la Banque de développement du Canada permettent d'investir dans des capacités d'IA et les mettre en œuvre.

Bien que la majorité de ces programmes soutiennent les entreprises désireuses d'adopter des capacités d'IA, ils ne fournissent pas de subventions directes pour les coûts de l'informatique en nuage de l'IIA ou pour l'achat de matériel.

Subventions et investissements internationaux pour les infrastructures pour l'IIA

Il faut également tenir compte de la résilience de la capacité d'IIA nationale en cas de perturbations du marché comme les tensions géopolitiques, l'insolvabilité des entreprises d'informatique en nuage ou les restrictions qui leur sont imposées. Plus explicitement, il existe un besoin croissant de protections pour l'écosystème de l'IA afin de garantir l'accès à l'informatique pour l'IA en cas de défaillance ou de perturbation du marché de l'IIA. En particulier, l'accès à l'infrastructure informatique en nuage basée aux États-Unis dépend d'un régime politique américain fiable en matière de protection des données et d'accès pour les entreprises non basées aux États-Unis.

Un certain nombre d'organismes nationaux ont mis en place une politique et une stratégie pour ces circonstances, énumérées ci-dessous.

Les États-Unis et le Japon ont tous deux promis d'appuyer la fabrication nationale de puces d'IIA, en réduisant la dépendance à l'égard de l'approvisionnement international de la chaîne d'approvisionnement de l'infrastructure pour l'IIA grâce à des investissements de plusieurs milliards de dollars dans la fabrication, la RetD et la main-d'œuvre qualifiée.^{50 51}

Bien que la majorité de ces programmes soutiennent les entreprises désireuses d'adopter des capacités d'IA, ils ne fournissent pas de subventions directes pour les coûts de l'informatique en nuage de l'IIA ou pour l'achat de matériel.

Dans le cadre du European High-Performance Computing Joint Undertaking (EuroHPC JU), l'Union européenne (UE) investira plus de 10 milliards de dollars dans l'ensemble de ses pays jusqu'en 2027. L'EuroHPC JU a alloué un total de 3 milliards de dollars à l'expansion de la capacité d'IIA parmi ses pays membres. L'UE a promis près de 700 millions de dollars par an pour appuyer les petites et moyennes entreprises d'IA en leur donnant accès à des sites informatiques à haute performance (IHP) et en finançant des ressources informatiques en nuage afin de garantir l'accès et l'évolutivité de leur écosystème d'IA.⁵² Sa branche d'investissement public, InvestEU, fournit plus de 1,6 milliard de dollars en appui financier pour l'incubation de jeunes entreprises, et plus de 700 millions de dollars aux entreprises établies pour qu'elles adoptent et déploient l'IA.⁵³

En Finlande, 20 % du superordinateur LUMI est mis à la disposition de l'industrie de l'IA, en partenariat avec des instituts de recherche. L'industrie peut également payer pour obtenir des ressources informatiques de haute performance, et les PME peuvent demander des subventions d'une valeur maximale de 117 000 dollars pour les utiliser à cette fin.⁵⁴ Le programme compense partiellement les coûts opérationnels et initiaux élevés des investissements publics dans les infrastructures d'IIA.

La France a promis 2,7 milliards de dollars pour appuyer son écosystème d'IA. Sa stratégie vise à appuyer les infrastructures d'informatique en nuage nationales et sécurisées pour protéger les données sensibles et à soutenir les fournisseurs français de services d'informatique en nuage pour un accès informatique de confiance.⁵⁵

En 2022, le Royaume-Uni a introduit un plan d'action pour l'IA prévoyant un appui de plus de 1,7 milliard de dollars pour le secteur, en complément des 3,8 milliards de dollars déjà investis.⁵⁶ Le Royaume-Uni a récemment lancé un appel d'offres pour des subventions d'infrastructure d'IIA d'un montant maximal de 900 millions de dollars à des organismes nationaux capables d'héberger et d'exploiter un minimum de 2 000 GPU, avec une mention particulière pour les entreprises en phase de démarrage et les entreprises de recherche.⁵⁷

L'Allemagne investit plus de 2,4 milliards de dollars dans l'IA pour l'exercice financier en cours et prévoit de déployer des sites IHP prêts pour l'IIA avec au moins 100 000 GPU par site. Le plan d'action en matière d'IA de son ministère fédéral de l'Éducation et de la Recherche (BMBWF) vise à améliorer l'accès au financement et aux infrastructures publiques pour les PME de son écosystème en 2024.⁵⁸

En Finlande, 20 % du superordinateur LUMI est mis à la disposition de l'industrie de l'IA, en partenariat avec des instituts de recherche. L'industrie peut également payer pour obtenir des ressources informatiques de haute performance, et les PME peuvent demander des subventions d'une valeur maximale de 117 000 dollars pour les utiliser à cette fin.

Tableau 4. Politiques nationales d'investissement dans l'IIA

Pays	Titre	Investissement total (en millions de dollars canadiens) ⁵⁹	Début de la période	Fin de la période	Type de politique	Critères
UE	High Performance Computing Joint Undertaking	10 000 \$	2022	2027	Subventionner et élargir l'accès à l'IIA	Jeunes entreprises, PME, écosystème de recherche
Finlande	LUMI Business Resource Program	S.O.	2021	2027	Subventionner et élargir l'accès à l'IIA	Industrie, écosystème de la recherche
États-Unis	CHIPS et Science Act	70 000 \$	2022		Appui à l'industrie des semi-conducteurs	Industrie, écosystème de la recherche
France	Stratégie nationale cloud	2 700 \$	2021	2025	Subventionner et élargir l'accès à l'IIA	Jeunes entreprises, PME, écosystème de recherche
Royaume-Uni	Plan d'action en matière d'IA	1 700 \$	2021	2031	Subventionner et élargir l'accès à l'IIA	Jeunes entreprises, PME, écosystème de recherche
Allemagne	Plan d'action en matière d'IA	2 400 \$	2023	2024	Subventionner et élargir l'accès à l'IIA	Industrie, écosystème de la recherche
Japon	Budget supplémentaire 2023-2024	18 000 \$	2023	2024	Appui à l'industrie des semi-conducteurs	Industrie


Infrastructure de l'IIA comme moyen d'améliorer la gouvernance de l'IA

Un développement important de l'infrastructure informatique pour l'IA appuie la résilience d'une autre manière. Bien que les mécanismes exacts ne soient pas encore tout à fait au point, l'infrastructure informatique pour l'IA peut être un vecteur de systèmes d'IA plus sûrs, mieux adaptés aux objectifs des politiques publiques et capables de soutenir la collaboration internationale dans le domaine de l'IA. Les auteurs de « Computing Power and the Governance of Artificial Intelligence » le disent clairement :

« Par rapport à d'autres intrants clés de l'IA (données et algorithmes), l'informatique pertinente pour l'IA est un point d'intervention particulièrement efficace : elle est détectable, excluable et quantifiable, et elle est produite par une chaîne d'approvisionnement extrêmement concentrée. Ces caractéristiques, ainsi que l'importance particulière de l'informatique pour les modèles d'IA de pointe, suggèrent que la gestion de l'informatique peut contribuer à la réalisation d'objectifs politiques communs, comme la garantie de la sécurité et de l'utilisation bénéfique de l'IA. »⁶⁰

Toutefois, pour que le Canada puisse participer plus pleinement aux conversations sur la gouvernance de l'IA par le biais de l'informatique pour l'IA, il doit disposer d'une plus grande part de l'infrastructure de l'informatique pour l'IA ou d'un meilleur accès à elle.

En résumé, les progrès en matière d'accès direct à l'IIA publique, d'appui direct aux entreprises et d'investissements dans l'acquisition et la production d'IA par les comparateurs internationaux creusent davantage l'écart entre le Canada et ses pairs en matière de capacité d'IIA.



En résumé, les progrès en matière d'accès direct à l'IIA publique, d'appui direct aux entreprises et d'investissements dans l'acquisition et la production d'IA par les comparateurs internationaux creusent davantage l'écart entre le Canada et ses pairs en matière de capacité d'IIA.

4

Principales leçons et scénarios politiques

Dans la section précédente, nous avons déterminé les enjeux clés suivants :

- **Capacité d'IIA :** Le Canada se trouve derrière tous les autres pays du G7 en matière d'infrastructures d'IIA. L'infrastructure d'IIA dont il dispose est insuffisante pour appuyer la recherche exploratoire. Moins de 10 % des infrastructures d'IIA situées sur le territoire national appartiennent au secteur privé. L'investissement dans la construction d'infrastructures nationales est relativement faible, et la chaîne d'approvisionnement pour appuyer de tels projets n'en est qu'à ses débuts.
- **Efficacité de l'IIA :** Le Canada se sert principalement de sa capacité d'IIA de recherche dans les domaines de la science biomédicale, de la science du climat et de l'ingénierie aérospatiale. Les entreprises privées doivent recourir à des solutions basées sur l'informatique en nuage pour accéder à l'IIA en dehors du Canada, et les pressions liées au prix diminuent leur utilisation.
- **Résilience de l'IIA :** La politique publique canadienne qui façonne la capacité à construire une IIA résiliente repose sur des programmes généraux qui encouragent l'adoption de technologies en général, et il n'existe pas de politiques adaptées qui ciblent spécifiquement l'accès à l'IIA. Cette situation contraste avec celle d'autres pays qui disposent de politiques spécifiquement ciblées.

Objectifs des politiques publiques et compromis

Nous supposons que l'objectif des politiques publiques du Canada est d'élargir l'accès à l'IIA pour réduire le déficit informatique pour l'IA, retenir les entreprises et les talents d'IA et promouvoir le développement et l'adoption de l'IA. Cet objectif peut promouvoir un objectif plus large de politique publique visant à accroître la productivité et la concurrence nationales. Comme Internet avant elle, la capacité d'IA pourrait alimenter un large éventail de domaines de la croissance économique qui nécessitent des solutions immédiates, à moyen et à long terme. L'IIA a le potentiel d'être traitée comme une infrastructure de type utilitaire.

Les politiques publiques doivent être conçues de manière à ce que l'investissement public profite au plus grand nombre d'entreprises et d'intervenants de l'écosystème de l'IA, que l'investissement public ne remplace pas simplement l'investissement privé et que toute participation publique soit bien gérée et s'adapte à l'évolution rapide de la politique et des circonstances du secteur.

Nous supposons que l'objectif des politiques publiques du Canada est d'élargir l'accès à l'IIA pour réduire le déficit informatique pour l'IA, retenir les entreprises et les talents d'IA et promouvoir le développement et l'adoption de l'IA.

Les options suivantes résument l'éventail limité de scénarios qui s'offrent au gouvernement du Canada. Les options sont classées par ordre décroissant de durée, compte tenu des besoins actuels importants de l'écosystème d'IA au Canada.

Scénario 1 : Centraliser et subventionner l'IIA par le biais de marchés publics fédéraux de la part de fournisseurs d'informatique en nuage

Proposition : On peut acheter des ressources informatiques pour l'IA auprès de services d'informatique en nuage d'entreprise. Pour aider à réduire le déficit informatique pour l'IA, ce scénario propose un effort national pour acheter des capacités à une échelle aussi grande que possible afin d'éliminer les obstacles immédiats à l'accès à l'IIA pour l'écosystème national de l'IA. L'ampleur des besoins immédiats en matière d'accès à l'IIA pourrait être évaluée par le biais d'une procédure d'appel d'offres et mise en œuvre par le biais de programmes d'attribution de plates-formes en nuage.

Un soutien direct aux coûts de l'informatique en nuage, similaire aux récents programmes administrés par la France, permettrait de satisfaire au besoin considérable en ressources informatiques endémique des innovateurs canadiens et des entreprises d'IA qui en sont à leurs débuts.

Idéalement, le gouvernement fédéral devrait acquérir des ressources d'IIA en nuage en bloc afin d'augmenter la probabilité d'accéder à une capacité rapide et évolutive au moindre coût, sur la base de la demande, du potentiel de croissance et des considérations d'équité pour l'IIA.

Mise en œuvre : Compte tenu de l'urgence de la question, le gouvernement devrait allouer des fonds et proposer des dispositions législatives au cours du processus budgétaire de 2024, avec l'intention de mettre en place un marché où les entreprises canadiennes pourraient s'approvisionner dès que possible. Bien que la demande totale des entreprises canadiennes et leur capacité à changer

de fournisseur à court terme soient actuellement inconnues, un programme d'achat de plusieurs centaines de millions par an permettrait de réduire les coûts globaux et d'améliorer l'accès à l'écosystème national de l'IA.⁶¹

Le coût fiscal pour le gouvernement dépendrait de l'ampleur de la subvention ou des remises accordées aux acheteurs et des types d'organismes admissibles (recherche, jeunes entreprises, PME, grandes entreprises).

Une autre option consiste à accorder des subventions directes en vertu de programmes de réduction pour les fournisseurs de services d'informatique dématérialisée avec des entreprises déjà enregistrées dans des programmes existants liés à la technologie. Cette subvention directe réduit la charge de l'allocation publique et de l'approvisionnement et peut être utilisée pour évaluer la demande totale du marché pour l'informatique pour l'IA en nuage par les jeunes entreprises, les PME et les chercheurs. L'adoption et le coût de l'approche des subventions directes seraient couverts par des ressources publiques, et le taux d'actualisation pourrait être ajusté en conséquence en fonction du type et de la taille de l'entreprise dans les futures itérations des programmes.

Pour : La centralisation des achats et les achats directs auprès des fournisseurs de services en nuage constituent la voie la plus immédiate pour améliorer l'accès à l'IIA canadienne à un prix abordable. Cette politique serait très bénéfique, en particulier pour les entreprises d'intelligence artificielle en phase de démarrage.

Contre : Ces subventions augmentent encore la dépendance du Canada à l'égard des États-Unis et des fournisseurs d'informatique en nuage. Même avec un pouvoir d'achat fédéral, la taille de notre marché pourrait ne pas nous permettre de réaliser des économies d'échelle suffisantes. Cette politique a un coût fiscal, bien que celui-ci puisse varier en fonction des décisions relatives au montant de la subvention disponible.

Scénario 2 : Travailler avec des partenaires commerciaux clés pour acheter conjointement l’IIA à grande échelle

Proposition : Le Canada peut acquérir des quantités encore plus importantes d’IIA et accroître son pouvoir de négociation en se joignant aux collaborations internationales existantes et, par le biais de la collaboration commerciale, en achetant avec une combinaison de partenaires européens, britanniques, japonais et autres. Le Canada est particulièrement bien placé pour mener une telle initiative en raison de son expertise en matière de commerce international, de diplomatie et d’intelligence artificielle. Ces partenariats permettraient de réduire les coûts d’acquisition de l’infrastructure de l’IIA et d’accéder à des services supplémentaires d’informatique en nuage à partir d’infrastructures d’informatique à haute performance publiques et privées existantes à l’échelle internationale, grâce à des accords d’achat, d’approvisionnement et de recherche conjoints.

Mise en œuvre : Le gouvernement canadien pourrait développer les accords récents avec le Royaume-Uni et le fabricant de puces Nvidia, ainsi que les travaux multilatéraux plus anciens avec le Partenariat mondial sur l’intelligence artificielle, entre autres, pour appuyer le développement d’une capacité nationale d’informatique pour l’AI par le biais de la délocalisation amicale et de l’élargissement de ses liens avec les différents acteurs de la chaîne d’approvisionnement de l’IA.^{62 63}

Le Canada conclut des négociations en 2024 pour adhérer au pilier II du programme de recherche et d’innovation Horizon Europe. En conséquence, le Canada pourrait envisager de se joindre à l’European High-Performance Computing Joint Undertaking (EuroHPC JU) et de se joindre à d’autres États non membres de l’EU, comme Israël et l’Islande. Le Canada devrait également appuyer et tirer parti des investissements massifs des États-Unis dans CHIPS et la Science Act ainsi que de sa relation unique avec les États-Unis pour négocier un meilleur accès à l’achat de matériel d’IIA à long terme.

Malgré certaines voix qui s’élèvent pour dire le contraire, le Canada ne devrait pas construire de capacité nationale de fabrication de puces (c’est-à-dire qu’il ne devrait pas construire de fonderies de puces).^{64 65 66} Il s’agirait d’une utilisation inefficace de l’argent fédéral et, du point de vue de la sécurité, du développement économique ou d’autres perspectives, le retour sur investissement ne serait pas satisfaisant. En revanche, l’achat proactif de matériel et d’infrastructures d’IIA à long terme pourrait permettre au Canada de disposer de chaînes d’approvisionnement durables par rapport à ses homologues internationaux. L’acquisition par plusieurs entreprises d’une infrastructure d’IA et de services en nuage peut nécessiter des obligations juridiques et contractuelles initiales plus complexes, mais elle réduira le risque d’une dépendance à long terme à l’égard d’une technologie et de services de qualité médiocre.⁶⁷ Bien que le gouvernement fédéral ait récemment fait allusion⁶⁸ à la nécessité de collaborer avec les alliés afin d’étudier les « possibilités communes » de fournir une plus grande puissance informatique, il convient de réfléchir à d’autres détails avant de mettre cette idée en œuvre.

Pour : Si les accords commerciaux et les initiatives multinationales prennent généralement du temps à se mettre en place, ils devraient être plus rapides que le renforcement des capacités nationales en matière d’AIIA et ils tirent parti des avantages uniques du Canada. L’échelle et le pouvoir d’achat revêtent la plus grande importance pour l’IIA en raison de leur impact sur le prix, après l’immédiateté de l’accès. Les achats multipartenaires robustes réduisent certains coûts pour l’écosystème national d’IIA. Ils pourraient également permettre aux Canadienne et Canadiens de disposer d’une plus grande capacité informatique générale, par rapport à d’autres pays qui disposeraient d’une plus grande capacité informatique au niveau national.

Contre : La manière dont les technologies sont régies à l’international évolue rapidement et la possibilité d’adhérer à des alliances pourrait diminuer au fil du temps. Cette approche peut également conduire le Canada à avoir des partenariats multiples avec des problèmes d’interopérabilité ou des performances moindres par dollar à long terme.

Scénario 3 : Créer une capacité nationale de superinformatique pour l'IA

Proposition : Le Canada achète du matériel informatique à haute performance préconstruit auprès de fournisseurs tiers d'informatique pour l'IA, avec des spécifications de performance similaires à celles du supercalculateur Olaf de la Corée du Sud, ce qui permettrait de déployer du matériel d'IIA supplémentaire pour l'écosystème d'IA sur un horizon de trois à cinq ans.⁶⁹ En outre, un programme public-privé d'allocation des ressources d'IIA, similaire au programme finlandais de recherche d'entreprises, pourrait être mis en place pour compenser l'investissement initial et résoudre les problèmes d'accès à l'IIA pour l'industrie.

En ce moment, l'infrastructure d'IIA nationale du Canada dispose de moins de la moitié de la capacité nécessaire pour atteindre la parité avec les concurrents internationaux. Par exemple, pour que le Canada triple sa capacité et ajoute 123 PFLOPS en infrastructure d'informatique spécifique à l'IIA Nvidia H100 et atteigne une capacité de performance d'IIA comparable au niveau international, nos estimations préliminaires suggèrent qu'il faudrait plus de 400 millions de dollars en coûts d'acquisition, d'intégration et d'exploitation du matériel spécifique à l'IIA.^{70 71 72 73} Yoshua Bengio a estimé que le coût d'une infrastructure d'informatique à haute performance suffisante dans son ensemble est d'environ un milliard de dollars.⁷⁴

Mise en œuvre : L'avenir incertain du rapport coût/performance de l'IIA complique la prévision de l'option claire pour les futurs investissements dans les infrastructures nationales. Les fournisseurs qui gèrent le paysage de plus en plus complexe de l'infrastructure informatique à grande échelle en conjonction avec un portefeuille d'entreprises de conception de puces d'IIA pourraient compenser le risque de pénuries potentielles dans la chaîne d'approvisionnement en IIA. Il serait prudent de conclure des accords pluriannuels d'infrastructure interopérable pour développer la capacité d'IIA nationale, qui pourrait être compensée par la revente ou la réaffectation de l'infrastructure

d'IIA traditionnelle, vieillissante et incompatible, pour satisfaire aux besoins de l'industrie et des établissements d'enseignement. Bien que l'intention immédiate soit de veiller à ce que l'infrastructure d'IIA spécialisée soit acquise à plus long terme, la nécessité de partenariats public-privé multi-entreprises devrait être envisagée pour garantir que des chaînes d'approvisionnement résilientes et à long terme soient établies pour les innovations imprévues dans l'écosystème d'IA au Canada.

Un partenariat public-privé pourrait permettre de développer à grande échelle l'informatique pour l'IA au niveau national. Le gouvernement fédéral pourrait aider à réunir et à joindre un consortium de grands fournisseurs de capitaux canadiens qui pourraient construire des capacités de superinformatique. La capacité pourrait être mise à la disposition de divers acteurs du secteur privé ainsi que du gouvernement fédéral à des taux concurrentiels sur le marché mondial. Le gouvernement fédéral pourrait alors utiliser son statut de membre du consortium pour aider à subventionner ou à rendre l'accès disponible à d'autres, y compris les jeunes entreprises, pour aider à atteindre les objectifs visant à garantir que davantage de talents et d'organismes en matière d'IA restent et se développent au Canada.

Pour : Ce scénario rapproche le Canada de l'autosuffisance, d'une plus grande parité de performances avec ses pairs internationaux et d'une plus grande souveraineté nationale en matière de capacité d'IA. L'allocation des ressources informatiques peut être plus rentable, étroitement surveillée et répartie de manière plus sûre entre les acteurs de la recherche et de l'industrie.

Contre : Il s'agit potentiellement du coût fiscal le plus élevé pour le gouvernement et il dépend de nombreux facteurs, notamment des pays étrangers et des pénuries d'approvisionnement. Pour cette option, l'acquisition et la mise en œuvre avec l'infrastructure informatique nationale existante se feraient à plus long terme qu'avec celle des ressources des fournisseurs de services en nuage.

Considérations politiques pour tous les scénarios

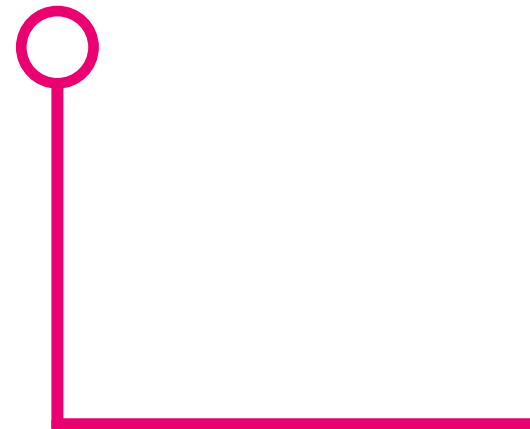
Les trois options ne s'excluent pas mutuellement et peuvent être mises en séquence. Par exemple, on peut poursuivre à court terme le scénario 1, Centraliser et subventionner, et le scénario 2, travailler avec des partenaires commerciaux clés, pour augmenter plus rapidement la capacité d'IIA et offrir une passerelle vers le « développement » national (scénario 3) de nouvelles capacités à plus long terme.

Les scénarios susmentionnés sont axés en particulier sur la capacité d'IIA et n'abordent pas les mécanismes d'attribution (c'est-à-dire l'efficacité de l'IIA) de la capacité nationale accrue de l'IIA. Étant donné que la majorité des entreprises privées accèdent aux ressources informatiques d'IA au niveau international, la compréhension de la demande globale de ces services nécessiterait une consultation considérable de l'écosystème d'IIA. Les mécanismes d'allocation doivent prendre en compte les perturbations potentielles du marché des investissements privés dans les infrastructures d'IIA. Une analyse séparée serait nécessaire pour identifier l'ampleur et la réaction dynamique aux programmes publics d'infrastructure et de subvention d'IIA. Cette analyse porterait spécifiquement sur la portée, les délais et les coûts qui découleraient dynamiquement de différents niveaux d'intervention publique sur le marché national de la prestation d'infrastructures d'IIA et de services en nuage.

Remédier au déficit informatique pour l'IA devrait s'harmoniser au positionnement du Canada comme que leader en matière d'innovation et d'adoption responsables de l'IA, y compris la Stratégie pancanadienne en matière d'IA, le processus législatif et réglementaire de la Loi sur *l'intelligence artificielle et les données* (LIAD), et d'autres initiatives.⁷⁵ Conformément à la pensée de certains de nos leaders en matière d'IA, l'investissement dans l'IIA peut être orienté vers des cas d'utilisation positifs, notamment dans le domaine de la santé et de la sécurité.⁷⁶

Remédier au déficit d'informatique pour l'IA doit tenir compte des implications en ce qui a trait à l'utilisation de l'énergie et aux objectifs climatiques. L'infrastructure d'IIA canadienne est actuellement deux fois moins efficace sur le plan énergétique que l'IA aux États-Unis.⁷⁷ Pourtant, l'électricité canadienne est essentiellement « verte »⁷⁸, le coût global de l'électricité est moins élevé que dans de nombreux autres pays, y compris les États-Unis et le Royaume-Uni,⁷⁹ et le climat plus froid⁸⁰ donne à l'IIA la possibilité d'être plus efficace dans ses besoins de refroidissement. Ces facteurs créent les conditions nécessaires à l'aspect concurrentiel des coûts pour attirer les investissements dans les infrastructures d'IIA au Canada, tout en s'harmonisant aux objectifs climatiques du pays.

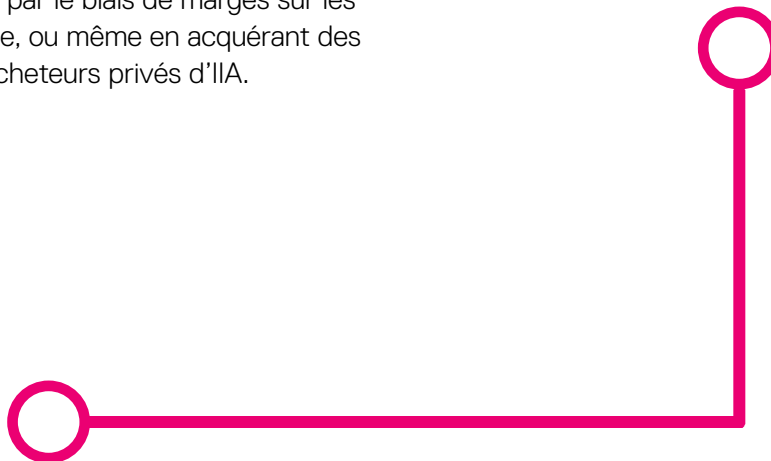
Ces options restent préliminaires et visent à alimenter des discussions rapides entre les acteurs de la politique et de l'écosystème d'IA. Des recherches, des analyses et des engagements industriels supplémentaires sont nécessaires pour évaluer la faisabilité et le rapport coût-efficacité de ces différentes options.



Considérations et risques liés à la conception de la politique

Lorsque le gouvernement envisage d'investir dans le secteur privé de l'IIA, il doit notamment tenir compte des éléments suivants :

- Demande insuffisante : L'un des principaux risques liés aux achats importants d'IIA par les pouvoirs publics est la sous-utilisation des capacités, l'augmentation des achats dépassant de loin la demande réelle actuelle en IIA. Étant donné que le mécanisme d'approvisionnement du scénario 1 repose sur des mécanismes non marchands, il convient de bien comprendre la demande avant de mettre en œuvre toute politique.
- L'éviction de l'investissement privé : Compte tenu de la très faible capacité nationale d'IIA accessible au secteur privé, tout programme d'achat/ subvention gouvernemental de grande envergure aura des répercussions de second ordre sur les prix, ce qui touchera les entreprises privées qui accèdent aux ressources d'IIA en dehors de ces programmes. Ces effets dynamiques d'éviction doivent être soigneusement modélisés et compris dans tous les scénarios.
- Mécanismes de marché innovants : Dans un avenir idéal, à moyen et long terme, pour les scénarios 2 et 3, les organismes qui déploient l'IIA nationale auraient une capacité suffisante pour utiliser les marchés afin de récupérer les coûts de cet investissement par le biais de marges sur les économies d'échelle, ou même en acquérant des parts auprès des acheteurs privés d'IIA.
- Tirer parti des modèles de marché technologiques antérieurs : Bien que très concentré, le marché du stockage en nuage est également devenu très dynamique, avec la possibilité de déplacer en temps réel les fonctions d'informatique et de stockage d'un fournisseur à l'autre. Tandis que le marché de l'IIA arrive à maturité, le Canada devra veiller à ce que son IIA puisse également rivaliser en termes de dynamisme.
- Inefficacité de la mise en œuvre : Il est particulièrement préoccupant que, si ces options sont retenues, la manière dont un programme particulier est mis en œuvre ait une incidence sur son utilisation et son adoption. Les décisions relatives à l'admissibilité, aux délais de traitement et aux seuils de subvention directe sont plus importantes que les ressources offertes par l'intervention publique. Cela peut être particulièrement important pour décider du véhicule de distribution approprié et de l'admissibilité pour le scénario 1, car la rapidité et l'accès sont les éléments les plus importants pour éliminer les obstacles immédiats au coût et à l'accès à l'IIA au Canada.



La suite

Autres considérations en matière de recherche :

Ayant commencé à impliquer les acteurs clés de l'écosystème d'IIA du Canada, the Dais s'engage à continuer à informer cette stratégie. En particulier, nous pensons pouvoir appuyer le développement de cette stratégie au cours de l'année à venir, par les moyens suivants :

- Organiser des tables rondes et engager une consultation approfondie et équilibrée (par exemple entre l'industrie, la société civile, le monde universitaire et les pouvoirs publics) couvrant l'écosystème d'IIA afin de comprendre l'ensemble des besoins en matière d'approvisionnement et d'exploitation pour l'IIA nationale
- Aider les décideurs politiques à comprendre les considérations climatiques et énergétiques des investissements supplémentaires dans l'infrastructure d'IIA, à savoir la capacité régionale, les contraintes électriques et l'interopérabilité de l'infrastructure d'informatique à haute performance existante
- Créer un cadre permettant d'estimer la demande globale et le coût total de l'informatique pour l'IA, étant donné que l'ampleur et la prestation de services (c'est-à-dire la préférence pour l'informatique en nuage ou l'infrastructure spécifique à l'IA) restent incertaines. Cela aidera à comprendre le montant approprié des achats ou des subventions du scénario 1.

En réalisant une analyse économique plus poussée de l'IIA nationale, the Dais peut fournir les preuves nécessaires aux décideurs pour mieux allouer les ressources en vue de développer et d'attirer les investissements en capital (fonds d'investissement et de pension, par exemple) pour l'expansion de l'infrastructure d'IIA nationale et les programmes d'appui.

5

Conclusion

Pour citer le ministre canadien de l'Innovation, des Sciences et du Développement économique (ISED) lors de l'annonce récente du partenariat Canada-Royaume-Uni en matière d'IA : « Nous avons des cerveaux. Il nous faut maintenant... le gros ordinateur. »⁸¹ La stratégie du Canada en matière d'IA est axée sur les talents, mais elle ne dispose pas de l'infrastructure nécessaire pour offrir des occasions aux entreprises en phase de démarrage, protéger les intérêts publics et exploiter le potentiel de l'IA pour une prospérité économique à long terme. Pour développer nos jeunes entreprises d'IA et commercialiser de manière responsable nos recherches dans ce domaine, nous devons collaborer avec d'autres pays et investir dans des capacités informatiques d'une qualité et d'une ampleur à la hauteur de ces ambitions, et ce, pour les Canadiennes et les Canadiens. Si le Canada veut être pris au sérieux comme leader dans l'écosystème international de l'IA, nous devons agir, car notre capacité actuelle ne satisfait pas aux exigences de la recherche appliquée, de l'innovation industrielle et de la gouvernance. En même temps, le Canada doit appuyer l'innovation responsable.

La capacité en informatique pour l'intelligence artificielle est le facteur le plus important qui manque à la croissance de l'écosystème canadien de l'intelligence artificielle. Si nous sommes d'accord avec les justifications pour investir davantage dans cet écosystème, à hauteur de centaines de millions ou de milliards de dollars – et les gains de productivité à eux seuls, s'ils sont distribués de manière appropriée, peuvent le justifier – alors nous devons agir dans les mois qui viennent, et ne pas attendre des années, pour négocier et mettre en place une stratégie d'IIA de classe mondiale.

Si le Canada veut être pris au sérieux comme leader dans l'écosystème international de l'IA, nous devons agir, car notre capacité actuelle ne satisfait pas aux exigences de la recherche appliquée, de l'innovation industrielle et de la gouvernance. En même temps, le Canada doit appuyer l'innovation responsable.

Endnotes

- ¹ « Real Talk: How Generative AI Could Close Canada's Productivity Gap and Reshape the Workplace-Lessons From the Innovation Economy », *Le Conference Board du Canada* (blogue), 20 février 2024, <https://www.conferenceboard.ca/product/real-talk/>.
- ² Amba Kak West et Sarah Myers, « The Problem With Public-Private Partnerships in AI », *Foreign Policy* (blogue), 22 février 2024, <https://foreignpolicy.com/2024/02/12/ai-public-private-partnerships-task-force-nairr/>.
- ³ « 2018 ACM A.M. Turing Award Laureates », Association for Computing Machinery, consulté le 21 février 2024, <https://awards.acm.org/about/2018-turing>.
- ⁴ « AICan : L'impact de la Stratégie pancanadienne en matière d'IA », CIFAR, consulté le 12 mars 2024, <https://cifar.ca/fr/ia/impact/>.
- ⁵ « Des compétences sur mesure : Évaluation des microcertifications chez les professionnels du secteur numérique », The Dais, 21 octobre 2023, <https://dais.ca/reports/built-to-scale-microcredentials-use-among-digital-professionals/>.
- ⁶ « U.S. AI Workforce. » Center for Security and Emerging Technology (blogue), consulté le 17 mars 2024, <https://cset.georgetown.edu/publication/u-s-ai-workforce/>.
- ⁷ Deloitte Canada, « Impact et Opportunités : L'écosystème de l'IA au Canada en 2023 », consulté le 21 février 2024.
- ⁸ « Future of Work Report: AI at Work », LinkedIn, consulté le 21 février 2024, <https://economicgraph.linkedin.com/research/future-of-work-report-ai>.
- ⁹ Crunchbase. « List of Top Canada Artificial Intelligence (AI) Companies - Crunchbase Hub Profile. » Consulté le 21 février 2024, <https://www.crunchbase.com/hub/canada-artificial-intelligence-companies>.
- ¹⁰ OCDE, « Données en temps réel OECD.AI », consulté le 17 mars 2024, <https://oecd.ai/fr/data>.
- ¹¹ Deloitte Canada, « Impact et Opportunités : L'écosystème de l'IA au Canada en 2023 », consulté le 21 février 2024, <https://www2.deloitte.com/ca/fr/pages/deloitte-analytics/articles/impact-and-opportunities-canadas-ai-ecosystem-2023.html>.
- ¹² Scale AI, « Portrait de l'IA d'ici », consulté le 21 février 2024, <https://www.scaleai.ca/fr/portraitia-2023/>.
- ¹³ « Automatisation à l'échelle nationale? Adoption de l'IA dans les entreprises canadiennes », The Dais, 14 août 2023, <https://dais.ca/reports/automation-nation-ai-adoption-in-canadian-businesses/>.
- ¹⁴ Scale AI, « Portrait de l'IA d'ici », consulté le 21 février 2024, <https://www.scaleai.ca/fr/portraitia-2023/>.
- ¹⁵ Guido Appenzeller, Matt Bornstein et Martin Casado, « Navigating the High Cost of AI Compute », Andreessen Horowitz, 27 avril 2023, <https://a16z.com/navigating-the-high-cost-of-ai-compute/>.
- ¹⁶ Murad Hemmadi, « Cloud Giants Ride Wave of AI Enthusiasm in Canada », *The Logic*, 16 janvier 2024, <https://thelogic.co/news/cloud-giants-ride-wave-of-ai-enthusiasm-in-canada/>.
- ¹⁷ Murad Hemmadi, « "To Compete, You Must Compute": Powering Canada's AI Surge », *The Logic*, 30 novembre 2023, <https://thelogic.co/news/special-report/to-compete-you-must-compute-powering-canadas-ai-surge/>.
- ¹⁸ OCDE, « A Blueprint for Building National Compute Capacity for Artificial Intelligence », Paris : OCDE, 28 février 2023, <https://doi.org/10.1787/876367e3-en>.
- ¹⁹ « Principes sur l'IA », OCDE, consulté le 18 mars 2024, <https://oecd.ai/fr/ai-principles>.
- ²⁰ Stuart Russell, Karine Perset et Marko Grobelnik, « Updates to the OECD's Definition of an AI System Explained », OECD.AI (blogue), 29 novembre 2023, <https://oecd.ai/en/wonk/ai-system-definition-update>.
- ²¹ OECD, « A Blueprint for Building National Compute Capacity for Artificial Intelligence », OECD Digital Economy Papers, no 350, 2023, OECD Publishing, <https://doi.org/10.1787/876367e3-en>.
- ²² Les unités centrales graphiques, ou GPU, sont des unités de traitement informatique spécialisées dans les calculs de traitement et d'affichage des graphiques informatiques, en particulier les graphiques complexes utilisés dans les jeux vidéo ou d'autres applications graphiques en 3D. Les calculs nécessaires au traitement des images de synthèse étant numériquement similaires aux calculs requis dans les modèles d'apprentissage automatique, de nombreuses entreprises spécialisées dans l'IA les ont préférées aux unités centrales de traitement, ou CPU. Les unités de traitement de tenseur (TPU), un produit plus récent, sont des unités de traitement informatique qui se spécialisent spécifiquement dans la formation de modèles d'apprentissage automatique.
- ²³ Jai Vipra et Sarah Myers West, « Computational Power and AI », AI Now Institute (blogue), 27 septembre 2023, <https://ainowinstitute.org/publication/policy/compute-and-ai>.
- ²⁴ OCDE, « A Blueprint for Building National Compute Capacity for Artificial Intelligence », Paris : OCDE, 28 février 2023, <https://doi.org/10.1787/876367e3-en>.
- ²⁵ Les PFLOPS, ou pétaflops, sont la mesure de la performance informatique totale selon la norme de référence LINPACK pour les infrastructures informatique à grande échelle. 1 PFLOP équivaut à 10¹⁵ (1 quadrillion) FLOPS.
- ²⁶ Pour tenir compte des différences de taille du marché économique et de population, nous calculons un ratio d'ajustement par habitant (<https://data.worldbank.org/indicator/SP.POP.TOTL?locations=CA-US-GB-FR-DE-JP-IT>) (en anglais) et PIB (<https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=CA-US-GB-FR-DE-JP-IT>) (en anglais) en utilisant le générateur de sous-liste de superordinateurs de TOP500 en date de novembre 2023, <https://www.top500.org/statistics/list/> (en anglais).
- ²⁷ Simon Nakonechny, « AI Pioneer Yoshua Bengio Urges Canada to Build \$1B Public Supercomputer », *CBC News*, 29 janvier 2024, <https://www.cbc.ca/news/canada/montreal/bengio-asks-canada-to-build-ai-supercomputer-1.7094858>.
- ²⁸ Chuan Li, « OpenAI's GPT-3 Language Model: A Technical Overview », *Lambda* (blogue), 3 juin 2020, <https://lambdalabs.com/blog/demystifying-gpt-3>.

- ²⁹ Vass Bednar, « Why Canada Needs a Publicly Owned Cloud », *Financial Post*, 17 janvier 2023, <https://financialpost.com/telecom/why-canada-needs-publicly-owned-cloud>.
- ³⁰ Accessible en cherchant « Canada » dans le générateur de sous-liste du Top 500 : <https://www.top500.org/statistics/sublist/>. RMax est le résultat d'un test d'étalement populaire, les scores les plus élevés indiquant une meilleure qualité. Rpeak divise le nombre total de PFlop par le nombre d'instructions pouvant être émises par seconde, et peut être interprété comme une mesure d'efficacité.
- ³¹ « Learn about AWS's Long-Term Commitment to Virginia. » 2023. US About Amazon. 7 juin 2023. <https://www.aboutamazon.com/news/aws/aws-commitment-to-virginia>.
- ³² Anita Balakrishnan, « Here's How Canada's Semiconductor Industry Stacks up », *The Logic*, 5 juillet 2023, <https://thelogic.co/news/special-report/heres-how-canadas-semiconductor-industry-stacks-up/>.
- ³³ Murad Hemmadi, « Tenstorrent Joins Japan's AI Compute Efforts », *The Logic*, 27 février 2024, <https://thelogic.co/briefing/tenstorrent-joins-japans-ai-compute-efforts/>.
- ³⁴ « Fact Sheet: CHIPS and Science Act Will Lower Costs, Create Jobs, Strengthen Supply Chains, and Counter China », The White House, 9 août 2022, <https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/>.
- ³⁵ About SciNet, « SciNet: Advanced Research Computing at the University of Toronto », consulté le 21 février 2024, <https://www.scinethpc.ca/about-scinet/>.
- ³⁶ « Systèmes nationaux – document de l'Alliance », consulté le 21 février 2024, https://docs.alliancecan.ca/wiki/National_systems/fr#grappes_de_calcul.
- ³⁷ Simon Nakonechny, « AI Pioneer Yoshua Bengio Urges Canada to Build \$1B Public Supercomputer », *CBC News*, 29 janvier 2024, <https://www.cbc.ca/news/canada/montreal/bengio-asks-canada-to-build-ai-supercomputer-1.7094858>.
- ³⁸ Murad Hemmadi, « “To Compete, You Must Compute” : Powering Canada's AI Surge », *The Logic*, 30 novembre 2023, <https://thelogic.co/news/special-report/to-compete-you-must-compute-powering-canadas-ai-surge/>.
- ³⁹ Scale AI. « Portrait de l'IA d'ici », consulté le 21 février 2024, <https://www.scaleai.ca/fr/portraitia-2023/>.
- ⁴⁰ Will Knight, « OpenAI's CEO Says the Age of Giant AI Models Is Already Over », *WIRED*, 17 avril 2023, <https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/>.
- ⁴¹ « FTC Launches Inquiry into Generative AI Investments and Partnerships », Federal Trade Commission, 25 janvier 2024, <https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships>.
- ⁴² Tom Krazit, « Do Big Cloud Companies Control AI Startups? », *Runtime*, 26 janvier 2024, <https://www.runtime.news/do-big-cloud-companies-control-ai-startups/>.
- ⁴³ Murad Hemmadi, « “To Compete, You Must Compute” : Powering Canada's AI Surge », *The Logic*, 30 novembre 2023, <https://thelogic.co/news/special-report/to-compete-you-must-compute-powering-canadas-ai-surge/>.
- ⁴⁴ « Résidence des données dans Azure », Microsoft Azure, consulté le 14 mars 2024, <https://azure.microsoft.com/fr-ca/explore/global-infrastructure/data-residency/>.
- ⁴⁵ « Microsoft Expands Digital Footprint in Quebec with USD \$500 million Investment in Infrastructure and Skilling Initiatives », Microsoft News Center Canada, 22 novembre 2023, <https://news.microsoft.com/en-ca/2023/11/22/microsoft-expands-digital-footprint-in-quebec-with-usd500-million-investment-in-infrastructure-and-skilling-initiatives/>.
- ⁴⁶ « Cloud locations - Regions & Zones », Google Cloud, consulté le 14 mars 2024, <https://cloud.google.com/about/locations>.
- ⁴⁷ Murad Hemmadi, « Cloud Giants Ride Wave of AI Enthusiasm in Canada », *The Logic*, 16 janvier 2024, <https://thelogic.co/news/cloud-giants-ride-wave-of-ai-enthusiasm-in-canada/>.
- ⁴⁸ « Canadian Cloud Hosting Services », Amazon Web Services, consulté le 14 mars 2024, <https://aws.amazon.com/local/canada/>.
- ⁴⁹ « Stratégie pancanadienne en matière d'intelligence artificielle », Innovation, Sciences et Développement économique Canada, 20 juillet 2022, <https://ised-isde.canada.ca/site/strategie-ia/fr>.
- ⁵⁰ Tetsushi Kajimoto et Sam Nussey, « Japan to Spend \$13 Bln for Chip Industry Support in Extra Budget », Reuters, 10 novembre 2023, <https://www.reuters.com/markets/asia/japan-allocate-13-bln-chip-industry-support-extra-budget-2023-11-10/>.
- ⁵¹ « Fact Sheet: CHIPS and Science Act Will Lower Costs, Create Jobs, Strengthen Supply Chains, and Counter China », The White House, 9 août 2022, <https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/>.
- ⁵² « AI Innovation Package to Support Artificial Intelligence Startups and SMEs Policy », OCDE, consulté le 28 février 2024, <https://oecd.ai/en/dashboards/policy-initiatives/http://aipo.oecd.org/2021-data-policy/initiatives-27588>.
- ⁵³ Commission européenne, « Communication sur le thème “Accroître les jeunes pousses et l'innovation dans le domaine de l'intelligence artificielle digne de confiance”, Commission européenne : Bâtir l'avenir numérique de l'Europe, 24 janvier 2024, <https://digital-strategy.ec.europa.eu/fr/library/communication-boosting-startups-and-innovation-trustworthy-artificial-intelligence>.
- ⁵⁴ « Solutions for companies – Implementation and Pricing », LUMI, consulté le 28 février 2024, <https://www.csc.fi/ratkaisut-yrityksille-laskentapalveluiden-kaytto>.
- ⁵⁵ « Stratégie nationale cloud : Lancement du plan industriel de soutien à la filière », Entreprises.Gouv.Fr., consulté le 8 février 2024, <https://www.entreprises.gouv.fr/fr/actualites/numerique/strategie-nationale-cloud-lancement-du-plan-industriel-de-soutien-la-filiere>.
- ⁵⁶ International Trade Administration, « United Kingdom Artificial Intelligence Market 2023 », 29 mars 2023, <https://www.trade.gov/market-intelligence/united-kingdom-artificial-intelligence-market-2023>.
- ⁵⁷ « Host Sites for the next Wave of UK Government AI Infrastructure », UK Research and Innovation, 24 janvier 2024, <https://www.ukri.org/opportunity/host-sites-for-the-next-wave-of-uk-government-ai-infrastructure/>.

- ⁵⁸ « BMBF Action Plan “Artificial Intelligence”, Bundesministerium Für Bildung Und Forschung, 7 novembre 2023, <https://www.bmbf.de/bmbf/de/forschung/digitale-wirtschaft-und-gesellschaft/kuenstliche-intelligenz/ki-aktionsplan.html>.
- ⁵⁹ Converti en dollars canadiens de mars 2024, la valeur originale est basée sur l’année d’annonce.
- ⁶⁰ Girish Sastry et coll., « Computing Power and the Governance of Artificial Intelligence », *Arxiv*, 13 février 2024, [arXiv:2402.08797](https://arxiv.org/abs/2402.08797).
- ⁶¹ En 2023, 180 entreprises d’IA canadiennes ont reçu plus de 2,5 milliards de dollars de financement en capital-risque, selon Crunchbase et l’OCDE; on estime que plus de la moitié du financement des entreprises de genAI va aux coûts de l’IIA. <https://oecd.ai/en/data?selectedArea=investments-in-ai-and-data&selectedVisualization=vc-investments-in-generative-ai-by-country>.
- ⁶² Canada, Affaires mondiales, « Protocole d’entente entre le ministère des Affaires étrangères, du Commerce et du Développement du Canada et le ministère des Sciences, de l’Innovation et de la Technologie du Royaume-Uni de Grande-Bretagne et d’Irlande du Nord pour la coopération concernant la recherche scientifique et l’innovation », 30 janvier 2024, https://www.deleguescommerciaux.gc.ca/innovators-innovateurs/mou_science_innovation_protocole_entente.aspx?lang=fra.
- ⁶³ Brian Caulfield, « Canada Partners With NVIDIA to Supercharge Computing Power », NVIDIA (blogue), 5 février 2024, <https://blogs.nvidia.com/blog/canada/>.
- ⁶⁴ « OpenAI’s Altman in Talks to Raise Funds for Chips, AI Initiative – WSJ », Reuters, 9 février 2024, <https://www.reuters.com/technology/openais-altman-talks-raise-funds-chips-ai-initiative-wsj-2024-02-09/>.
- ⁶⁵ Anton Shilov, « Legendary Chip Architect Jim Keller Responds to Sam Altman’s Plan to Raise \$7 Trillion to Make AI Chips – “I Can Do It for Less than \$1 Trillion” », *Tom’s Hardware*, 17 février 2024, <https://www.tomshardware.com/tech-industry/artificial-intelligence/jim-keller-responds-to-sam-altmans-plan-to-raise-dollar7-billion-to-make-ai-chips>.
- ⁶⁶ Anton Shilov, « Nvidia CEO Jensen Huang Says \$7 Trillion Isn’t Needed for AI – Cites 1 Million-Fold Improvement in AI Performance in the Last Ten Years », *Tom’s Hardware*, 14 février 2024, <https://www.tomshardware.com/pc-components/gpus/nvidia-ceo-jensen-huang-says-dollar7-trillion-isnt-needed-for-ai-cites-1-million-fold-improvement-in-ai-performance-in-the-last-ten-years>.
- ⁶⁷ Tandis que Nvidia détient une part dominante du marché du matériel d’IIA, AMD, Intel et potentiellement Apple rattrapent leur retard en matière de performances d’IIA. Voir : <https://www.top500.org/statistics/list/>. En novembre 2023, les coprocesseurs Nvidia détiendront plus de 25 % de la liste des 500 meilleurs superordinateurs, contre moins de 3 % pour AMD et 2 % pour Intel.
- ⁶⁸ Murad Hemmadi, « G7 Nations Exploring Cooperating on AI Compute: Champagne », *The Logic*, 15 mars 2024.
- ⁶⁹ « Olaf – Lenovo ThinkSystem SR675 V3, AMD EPYC 9334 32C 2,7 GHz, NVIDIA H100, Infiniband NDR 400 » TOP500, consulté le 21 février 2024, <https://www.top500.org/system/180180/>.
- ⁷⁰ En termes réels, le coût des cartes de traitement graphique NVIDIA H100 est d’environ 45 000 dollars, et elles sont capables d’effectuer 34 TFLOPS sur la base LINPACK. NVIDIA. « NVIDIA H100 Tensor Core GPU Datasheet », consulté le 21 février 2024, <https://resources.nvidia.com/en-us-tensor-core/nvidia-tensor-core-gpu-datasheet>.
- ⁷¹ « NVIDIA H100 – GPU Computing Processor – NVIDIA H100 Tensor Core – 80 GB – 900-21010-0000-000 – Graphic Cards », CDW.ca, consulté le 21 février 2024, <https://www.cdw.ca/product/nvidia-h100-gpu-computing-processor-nvidia-h100-tensor-core-80-gb/7354651>.
- ⁷² En supposant que l’infrastructure nationale dispose des emplacements PCIe et de la capacité d’alimentation nécessaires pour ajouter plus de 3 600 cartes accélératrices Nvidia H100 à technologie PCIe..
- ⁷³ « NVIDIA H100 – GPU Computing Processor », SHI.ca, consulté le 21 février 2024, <https://www.shi.ca/Product/45671009/NVIDIA-H100-GPU-computing-processor>.
- ⁷⁴ Simon Nakonechny, « AI Pioneer Yoshua Bengio Urges Canada to Build \$1B Public Supercomputer », *CBC News*, 29 janvier 2024, <https://www.cbc.ca/news/canada/montreal/bengio-asks-canada-to-build-ai-supercomputer-1.7094858>.
- ⁷⁵ « Submission on the Proposed Artificial Intelligence and Data Act », The Dais, 13 novembre 2023, <https://dais.ca/reports/submission-on-the-proposed-artificial-intelligence-and-data-act/>.
- ⁷⁶ Comme il l’a indiqué dans un récent article plaidant en faveur d’une plus grande capacité d’IIA au Canada, Yoshua Bengio souhaiterait que cette catégorie de machines soit construite au Canada et financée par les gouvernements, afin que les entités publiques disposent de la puissance de feu numérique nécessaire pour suivre les géants privés de la technologie qu’elles seront chargées de surveiller ou de réglementer. « Je pense que le gouvernement devra comprendre à un moment donné, et j’espère le plus tôt possible, qu’il est important pour lui d’avoir cette force », a déclaré Bengio. Simon Nakonechny, « AI Pioneer Yoshua Bengio Urges Canada to Build \$1B Public Supercomputer », *CBC News*, 29 janvier 2024, <https://www.cbc.ca/news/canada/montreal/bengio-asks-canada-to-build-ai-supercomputer-1.7094858>.
- ⁷⁷ Si l’on considère la performance en fonction des kilowatts nécessaires pour alimenter l’infrastructure d’IIA à usage public, le nombre de PFLOPS par kilowatt au Canada est de 0,007, tandis qu’il est de 0,016 PFLOPS par kilowatt aux États-Unis. Cela signifie que l’infrastructure d’IIA actuelle à usage public du Canada est efficace du point de vue énergétique à moins de la moitié de celle de l’infrastructure informatique à usage public à plus grande échelle des États-Unis..
- ⁷⁸ Nour Abdelaal, Adams Aghmien et André Côté, « Clean Connection: How Digitization Can Support Canada’s Path to Net-Zero », The Dais, 23 juin 2023, <https://dais.ca/reports/clean-connection-how-digitization-can-support-canadas-path-to-net-zero/>.
- ⁷⁹ « Electricity Prices around the World », GlobalPetrolPrices.com, consulté le 12 mars 2024, <https://www.globalpetrolprices.com/electricity-prices/>.
- ⁸⁰ « Climate Comparison: Canada / United States », WorldData.info, consulté le 12 mars 2024, <https://www.worlddata.info/climate-comparison.php?r1=canada&r2=usa>.
- ⁸¹ Murad Hemmadi, « Canada and U.K. Are Allies, Not Rivals, on AI: Champagne and Donelan », *The Logic*, 31 janvier 2024, <https://thelogic.co/news/canada-and-u-k-are-allies-not-rivals-on-ai-champagne-and-donelan/>.