

The background of the top half of the image is a complex, golden-brown circuit board pattern. In the center, a white silhouette of a person is captured mid-jump, clearing a bar chart. The bar chart consists of four blue, 3D rectangular blocks of increasing height from left to right. The person's arms are raised and legs are extended, suggesting a leap of faith or achievement.

Entre progrès et défis

Le comité d'experts sur l'intelligence artificielle
en sciences et en génie



CCA | CAC

Entre progrès et défis

Le comité d'experts sur l'intelligence artificielle
en sciences et en génie



Conseil des académies canadiennes 180, rue Elgin, bureau 1401, Ottawa (Ontario) Canada K2P 2K3

Le projet sur lequel porte ce rapport a été entrepris avec l'approbation du conseil d'administration du Conseil des académies canadiennes (CAC). Les membres du conseil d'administration sont issus de la Société royale du Canada (SRC), de l'Académie canadienne du génie (ACG) et de l'Académie canadienne des sciences de la santé (ACSS), ainsi que du grand public. Les membres du comité d'experts responsables du rapport ont été choisis par le CAC en raison de leurs compétences particulières et dans le but d'obtenir un éventail équilibré de points de vue.

Ce rapport a été rédigé en réponse à une demande du Conseil national de recherches Canada et de l'Institut canadien de recherches avancées, des Instituts de recherche en santé du Canada, du Conseil national de recherches en sciences naturelles et en génie du Canada et du Conseil de recherche en sciences humaines. Les opinions, constatations et conclusions présentées dans cette publication sont celles des auteurs, soit les membres du Comité d'experts sur l'intelligence artificielle en sciences et en génie, et ne reflètent pas nécessairement le point de vue des organisations auxquelles ils sont affiliés ou dans lesquelles ils travaillent ou du commanditaire et des co-commanditaires.

Bibliothèque et Archives Canada

ISBN: 978-1-990592-12-6 (livre)

978-1-990592-13-3 (livre électronique)

Ce rapport devrait être cité comme suit :

Conseil des académies canadiennes, 2022. *Entre progrès et défis*. Ottawa, ON, Comité d'experts sur l'intelligence artificielle en sciences et en génie, Conseil des académies canadiennes.

Avis de non-responsabilité

Au meilleur de la connaissance du CAC, les données et les informations tirées d'Internet qui figurent dans ce rapport étaient exactes à la date de publication du rapport. En raison de la nature dynamique d'Internet, des ressources gratuites et accessibles au public peuvent subséquemment faire l'objet de restrictions ou de frais d'accès, et l'emplacement des éléments d'information peut changer lorsque les menus et les pages Web sont modifiés.



© 2022 Conseil des académies canadiennes
Imprimé à Ottawa, Canada



Ce projet a été rendu possible grâce au
soutien du gouvernement du Canada

Le comité d'experts sur l'intelligence artificielle en sciences et en génie aimerait remercier les Inuits, les Métis et les Premières Nations d'avoir assuré l'intendance du territoire qu'on appelle aujourd'hui le Canada.


Le Conseil des académies canadiennes (CAC) reconnaît que ses bureaux d'Ottawa sont situés sur le territoire ancestral non cédé et non abandonné de la Nation algonquine Anishinaabe, qui a toujours veillé à la terre, à l'eau et à l'air de ce territoire et continue à le faire aujourd'hui. Bien que les bureaux du CAC se trouvent à cet endroit, ses travaux en faveur de la prise de décision éclairée par des données probantes peuvent avoir des bienfaits plus étendus et contribuer, on l'espère, à l'action collective contre les iniquités et les injustices de longue date dont sont victimes les peuples autochtones. Nous sommes déterminés à tirer parti d'un éventail de connaissances et d'expériences afin de guider les politiques qui bâtiront une société plus forte, plus équitable et plus juste.

Le Conseil des académies canadiennes

Le Conseil des académies canadiennes (CAC) est un organisme indépendant à but non lucratif qui réalise des évaluations spécialisées indépendantes, étayées scientifiquement et faisant autorité, dans le but d'éclairer l'élaboration de politiques publiques au Canada. Dirigé par un conseil d'administration et conseillé par un comité consultatif scientifique, le CAC a pour champ d'action la science au sens large, ce qui englobe les sciences naturelles, les sciences humaines et sociales, les sciences de la santé, le génie et les lettres. Les évaluations du CAC sont effectuées par des comités pluridisciplinaires indépendants d'experts provenant du Canada et de l'étranger. Ces évaluations visent à cerner des problèmes nouveaux, des lacunes de nos connaissances, les atouts du Canada, ainsi que les tendances et les pratiques internationales. Ces études fournissent aux décideurs gouvernementaux, aux universitaires et aux parties prenantes l'information de grande qualité dont ils ont besoin pour élaborer des politiques publiques éclairées et innovatrices.

Tous les rapports d'évaluation du CAC sont soumis à un examen formel. Ils sont publiés et mis à la disposition du public sans frais. Des fondations, des organisations non gouvernementales, le secteur privé et tout ordre de gouvernement peuvent soumettre au CAC des questions susceptibles de faire l'objet d'une évaluation.

www.rapports-cac.ca

 @cca_reports

Les Académies

Le CAC est soutenu par ses trois académies fondatrices :

La Société royale du Canada (SRC)

Fondée en 1882, la SRC comprend l'Académie des arts, des lettres et des sciences, ainsi que le tout premier organisme canadien de reconnaissance multidisciplinaire destiné à la nouvelle génération d'intellectuels canadiens : le Collège de nouveaux chercheurs et créateurs en art et en science. Sa mission consiste à reconnaître les plus éminents intellectuels, chercheurs et créateurs, à conseiller les gouvernements et les organisations, et à favoriser l'avancement du savoir et de l'innovation au Canada avec d'autres académies nationales partout dans le monde.

L'Académie canadienne du génie (ACG)

L'ACG est l'organisme national par l'entremise duquel les ingénieurs les plus chevronnés et émérites du Canada offrent des conseils stratégiques sur des enjeux d'importance primordiale pour le pays. L'ACG est un organisme indépendant, autonome et à but non lucratif qui a été fondé en 1987. Les Fellows de l'ACG sont nommés et élus par leurs pairs, en fonction de leurs réalisations exceptionnelles et de leurs longs états de service au sein de la profession d'ingénieur. Les Fellows de l'ACG s'engagent à faire en sorte que l'expertise du Canada en ingénierie soit mise à contribution pour le plus grand bien de tous les Canadiens et de toutes les Canadiennes.

L'Académie canadienne des sciences de la santé (ACSS)

L'ACSS reconnaît l'excellence dans les sciences de la santé en nommant ses membres en fonction de leurs réalisations exceptionnelles dans les disciplines universitaires des sciences de la santé au Canada et de leur volonté de servir le public canadien. L'ACSS fournit des évaluations opportunes, factuelles et impartiales sur des sujets qui touchent la santé de la population canadienne, et recommande des solutions stratégiques et réalisables. Fondée en 2004, l'ACSS nomme de nouveaux membres sur une base annuelle. L'ACSS est dirigée par un conseil d'administration constitué de volontaires et par un comité de direction.

Comité d'experts sur l'intelligence artificielle en sciences et en génie

Guidé par son comité consultatif scientifique, son conseil d'administration et ses académies fondatrices, le CAC a constitué le comité d'experts sur l'intelligence artificielle en sciences et en génie pour mener à bien ce projet. Chacun des membres de ce comité a été choisi pour son expertise, son expérience et son leadership éprouvé dans des domaines pertinents pour le projet.

Teresa Scassa (présidente), titulaire de la chaire de recherche du Canada en politiques et droit de l'information, Faculté de droit de l'Université d'Ottawa (Ottawa, Ont.)

Julien Billot, président-directeur général, Scale AI (Montréal, Qc)

Wendy Hui Kyong Chun, titulaire de la chaire de recherche Canada 150 en nouveaux médias et professeure en communications, Université Simon Fraser (Burnaby, C.-B.)

B. Courtney Doagoo, gestionnaire principale, Services-conseils — Gestion des risques, KPMG LLP; chercheuse invitée en intelligence artificielle et société, Centre de recherche en droit, technologie et société, Université d'Ottawa (Toronto, Ont.)

Abhishek Gupta, fondateur et chercheur principal, Montreal AI Ethics Institute (Montréal, Qc)

Richard Isnor, vice-président associé, recherche et études supérieures, Université St. Francis Xavier (Antigonish, N.-É.)

Ross D. King, professeur, Université de technologie Chalmers (Göteborg, Suède); professeur, Université de Cambridge (Cambridge, Royaume-Uni)

Sabina Leonelli, professeure en philosophie et en histoire des sciences et directrice d'Egenis, Université d'Exeter (Exeter, Royaume-Uni); fellow de la Wissenschaftskolleg zu Berlin (Berlin, Allemagne)

Raymond J. Spiteri, professeur, Département d'informatique, Université de la Saskatchewan (Saskatoon, Sask.)

Le CAC remercie également **Marc-Antoine Dilhac**, professeur, Département de philosophie, Université de Montréal; titulaire de chaire en IA Canada-CIFAR; directeur, Algora Lab; coprésident de Délibérations, Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA) (Montréal, Qc), pour sa contribution.

Message du président-directeur général

L'intelligence artificielle (IA) continue à capter l'attention du monde, en grande partie parce qu'elle peut effectuer des activités qu'auparavant seuls les humains pouvaient accomplir. À ce jour, elle a été déployée aux côtés de pratiques de longue date de conception et de découverte pour aider les chercheurs à analyser et interpréter les données. Elle est par exemple employée pour prédire la structure des protéines, pour définir les composés chimiques en recherche biomédicale, pour sélectionner les biomarqueurs préférés pour le développement de médicaments potentiels et pour faire le suivi de la biodiversité des insectes.

Étant donné la rapidité avec laquelle l'IA continue à se développer, elle commencera très bientôt à jouer un rôle plus important dans la conception et la découverte en sciences et en génie. On prévoit qu'elle sera utilisée pour concevoir des hypothèses et des expériences scientifiques novatrices, et pour créer de nouveaux processus de conception en génie, en reposant de moins en moins sur la programmation humaine.

Pour réaliser les promesses et les bénéfices potentiels de l'IA, il faudra surmonter les défis réels et imminents concernant les biais possibles, que ce soit de la part des gens qui les construisent ou des établissements et des gouvernements qui élaborent les politiques pour la régir. Garantir l'utilisation responsable de l'IA peut stimuler l'innovation et améliorer la connaissance scientifique, mais il y aura des coûts pour cela.

Reconnaissant la grande complexité de l'utilisation de l'IA dans l'ensemble des disciplines, le Conseil national de recherches Canada et les autres commanditaires ont demandé au CAC d'examiner ses conséquences légales, réglementaires, éthiques, sociales et politiques.

Entre progrès et défis étudie les possibilités, les défis et les conséquences du déploiement des technologies d'IA pour permettre la conception de la recherche et la découverte en sciences et en génie. Ce rapport cerne les acteurs dont les décisions détermineront la façon dont ces défis seront surmontés et comment les divers champs et secteurs pourraient intégrer l'IA dans leur pratique.

Le comité d'experts, présidé par Teresa Scassa, comprenait des membres possédant une expertise en droit, en éthique, en sciences humaines, en sciences appliquées, en industrie et en politique. Comme de nombreuses évaluations récentes effectuées par le CAC pendant la COVID 19, le comité s'est réuni virtuellement du début à la fin. Nous remercions ses membres pour le temps, l'énergie et l'expertise qu'ils ont consacrés à l'exercice. Je remercie également le conseil d'administration et le comité consultatif scientifique du CAC et ses trois académies fondatrices, soit la Société royale du Canada, l'Académie canadienne du génie et l'Académie canadienne des sciences de la santé, pour les conseils et la supervision qu'ils ont fournis durant le processus.



Eric M. Meslin, Ph. D, MSRC, MACSS

Président-directeur général, Conseil des académies canadiennes

Message de la présidente

Les progrès en matière d'intelligence artificielle (IA) ont le potentiel de transformer la nature de la recherche scientifique et à mener à d'importantes innovations en génie. En tant qu'outil de recherche de plus en plus utilisé par les gens à beaucoup plus de stades de la conception et de la découverte, l'IA transforme les bases épistémiques des sciences et du génie. Mais l'IA se heurte à des écueils qui ont un important effet dans l'esprit des gens. Si elle n'est pas utilisée de façon responsable, celle-ci pourrait perpétuer les biais humains et exacerber les iniquités dans le système de recherche et dans la société plus généralement.

La prise en considération des conséquences sociales et éthiques de l'IA en sciences et en génie, des premières étapes du développement jusqu'au déploiement, sera cruciale à son usage réfléchi et responsable. Afin de pouvoir réaliser les bénéfices de l'IA, il sera également essentiel d'établir des mécanismes solides et transparents faisant en sorte que les résultats qu'elle génère sont exacts, reproductibles et explicables. Le succès reposera sur une meilleure collaboration entre les disciplines.

Au Canada, l'écosystème d'IA s'est concentré sur une croissance verticale, mais il devra croître horizontalement au-delà de ses forces actuelles et transcender les frontières physiques, disciplinaires et sectorielles pour maximiser les occasions de conception et de découverte en sciences et en génie.

Les cadres légaux et réglementaires actuels ont du mal avec plusieurs nouveaux défis découlant de l'utilisation d'un système d'IA à des fins prédictives ou décisionnelles, ce qui augmente les risques pour la confiance et la reddition de comptes. Les gouvernements et les décideurs devront déterminer s'il faut adapter les cadres actuels aux systèmes d'IA ou s'ils doivent élaborer de nouveaux cadres qui s'attaquent aux incertitudes entourant l'IA. Parce que de nombreuses parties prenantes contribuent au développement et au déploiement de l'IA en sciences et en génie, un défi crucial pour les décideurs sera non seulement de concevoir de nouvelles politiques, mais également de les coordonner entre des secteurs disparates.

L'intégration des connaissances et des compétences en sciences sociales, en sciences humaines et en sciences de la santé aidera à mieux comprendre l'IA pour les sciences et le génie et aura des répercussions en dehors du laboratoire. Fait important, la collaboration transdisciplinaire aidera également à régler les problèmes de longue date en matière d'équité, de diversité et d'inclusion associés à l'utilisation de l'IA dans le système de recherche canadien.

Agir comme présidente de ce comité d'experts a été un plaisir. J'aimerais remercier mes collègues membres du comité pour leur contribution et leur engagement tout au long du processus ainsi qu'au personnel du CAC pour son soutien constant et ses conseils. Enfin, je tiens à remercier les commanditaires de nous avoir soumis cette question est d'avoir rendu notre travail possible.

A handwritten signature in black ink, appearing to read 'Teresa Scassa', followed by three horizontal dashes.

Teresa Scassa, J.S.D.

Présidente, comité d'experts sur l'intelligence artificielle en sciences et en génie

Équipe de projet du Conseil des académies canadiennes

Équipe d'évaluation : **Jean Woo**, directrice de projet
Alexei Halpin, associé de recherche
Matthew Ivanowich, associé de recherche
Weronika Zych, coordinatrice de projet principale
Karl Guebert, consultant en recherche
Tijs Creutzberg, directeur des évaluations

Avec la participation de :

Conception	gordongroup
Révision	Jody Cooper Lisa Muirhead, Firefly Communications
Traduction, En-Fr	François Abraham, traducteur agréé, Communications Léon inc.

Examen du rapport

La version préliminaire de ce rapport a été revue par des examinateurs choisis par le CAC pour la diversité de leurs points de vue et de leurs domaines d'expertise. Ces examinateurs ont évalué l'objectivité et la qualité du rapport. Le comité a étudié intégralement leurs observations confidentielles et a intégré bon nombre de leurs suggestions. Le CAC ne leur a pas demandé de cautionner les conclusions du rapport et ils n'ont pas vu la version finale avant publication. La responsabilité du contenu final de ce rapport incombe entièrement au comité d'experts qui l'a rédigé et au CAC.

Le CAC tient à remercier les personnes suivantes pour leur examen du présent rapport :

Steven Berg, président et chef de la direction, Aquanty Inc. (Waterloo, Ont.)

Ignacio Cofone, professeur adjoint, Département de droit, et titulaire de la chaire de recherche du Canada sur le droit de l'intelligence artificielle et la gouvernance des données, Université McGill (Montréal, Qc)

Jason Edward Lewis, MSRC, professeur titulaire, Département de design et d'arts numériques, et titulaire de la chaire de recherche sur les médias numériques et l'imaginaire futuriste autochtone, Université Concordia; codirecteur, Indigenous Futures Research Centre (Montréal, Qc)

Fred Gault, professeur honoraire, Université des Nations unies — Maastricht Economic and Social Research Institute on Innovation and Technology (Maastricht, Pays-Bas); professeur extraordinaire, Université de technologie Tshwane (Pretoria, Afrique du Sud)

Maya Medeiros, associée, avocate et agente de brevets, Norton Rose Fulbright Canada LLP (Vancouver, C.-B.)

Emanuele Ratti, Institute of Philosophy and Scientific Method, Université Johannes Kepler (Linz, Autriche); Département des arts et des sciences humaines, Technion, Israel Institute of Technology (Haïfa, Israël)

Janna Rosales, professeure agrégée (enseignement), Faculté de génie et de sciences appliquées, Université Memorial (St. John's, T.-N.-L.)

L'examen du rapport a été supervisé, au nom du conseil d'administration et du comité consultatif scientifique du CAC, par **Nicole A. Poirier, FACG**, présidente, KoanTeknico Solutions Inc. Son rôle était de veiller à ce que le comité d'experts prenne en considération de façon entière et équitable les avis des examinateurs. Le conseil d'administration du CAC n'autorise la publication d'un rapport de comité d'experts qu'une fois que la personne chargée de superviser l'examen du rapport par les pairs a confirmé que celui-ci satisfait bien aux exigences du CAC. Le CAC remercie madame Poirier d'avoir supervisé consciencieusement l'examen du rapport.

Sommaire

L'intelligence artificielle (IA) favorise et améliore la découverte scientifique dans une variété de domaines. Elle a le potentiel de stimuler la recherche scientifique en permettant la génération d'hypothèses, la conception d'expériences, l'expérimentation, l'interprétation et l'analyse automatiques. D'un point de vue technique, le développement et le déploiement de systèmes d'IA sont favorisés par les progrès algorithmiques ainsi que par l'accès à un vaste et croissant volume de données. Étant donné la base scientifique des innovations qui deviennent ensuite des produits et des services largement utilisés, les conséquences de l'application de l'IA à la découverte scientifique seront inévitablement importantes.

Conscient des possibilités, des défis et des conséquences de l'application de l'IA aux sciences et au génie, le Conseil national de recherches Canada (NRC), avec l'appui de l'Institut canadien de recherches avancées (CIFAR), des Instituts de recherche en santé du Canada (IRSC), du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et du Conseil de recherches en sciences humaines (CRSH), a demandé au Conseil des académies canadiennes (CAC) de former un comité d'experts pour répondre à la question suivante :



Quels sont les défis légaux/réglementaires, éthiques, sociaux et politiques du déploiement des technologies d'IA pour permettre la conception de la recherche et la découverte en sciences et en génie au Canada?

Le CAC a donc constitué un comité multidisciplinaire et multisectoriel d'experts en droit, en éthique, en sciences humaines, en sciences appliquées, en industrie et en politique, qui s'est réuni virtuellement cinq fois entre 2021 et 2022 pour évaluer les données probantes et répondre à son mandat.

Constatations du rapport

Les capacités des systèmes d'IA se sont accrues et peuvent être appliquées à une variété croissante de tâches en sciences et en génie

Les systèmes d'IA sont depuis longtemps utilisés en sciences et en génie, en particulier comme outils d'analyse des données. Cependant, l'IA est aujourd'hui également appliquée à des tâches situées plus en amont dans la conception et la découverte, en raison des progrès technologiques dans les logiciels et dans le matériel, d'une croissance soutenue de la disponibilité des données et de l'émergence d'industries qui placent ces avancées au cœur de leurs activités. Les outils de recherche en IA non seulement améliorent l'analyse des résultats de recherche, mais leur capacité à reconnaître et prédire les modèles peut servir à déceler de nouveaux domaines de recherche et développement (R-D) pour les scientifiques et les ingénieurs. Il est possible d'étudier plus en profondeur des systèmes et des interactions complexes au moyen d'outils d'IA qui rendent les gros ensembles de données plus gérables. En fait, les applications d'IA établissent de nouveaux paradigmes dans certains secteurs, comme la découverte de médicaments, et améliorent les méthodes dans d'autres, comme dans le développement logiciel. À l'avenir, les systèmes d'IA pourraient jouer un plus grand rôle dans l'organisation de la recherche ou dans la proposition de concepts. Cette perturbation potentielle s'étend aux tâches cruciales dans divers flux en sciences et en génie, de la génération d'hypothèses à l'interprétation et à l'analyse. Ces tâches pourraient être exécutées de façon autonome par des systèmes d'IA à la suite d'avancées en robotique et dans les interfaces entre les machines et l'humain. Bien que ce scénario dépasse de beaucoup les capacités dont disposent les systèmes d'IA d'aujourd'hui — qui affichent encore de nombreuses limitations — les tendances donnent à penser que peu de champs seront laissés intacts par l'IA, ce qui justifie une préparation de la part de l'écosystème de R-D.

L'utilisation accrue de l'IA en sciences et en génie crée de nouveaux défis épistémiques, méthodologiques et éthiques pour les chercheurs

Bien que l'IA ait le potentiel d'améliorer la reproductibilité en sciences, elle est pour le moment gênée par le manque de transparence dans le partage de code et de données qui, en fin de compte, mine la confiance et l'exactitude des résultats. Des efforts sont en cours pour promouvoir la transparence dans les modèles et l'étalonnage par le biais d'ensembles de données partagés et par l'établissement d'exigences dans les publications et les conférences. La création de normes relatives aux données et aux métadonnées peut faciliter l'amélioration de la transparence dans ce domaine, tout comme le soutien aux données ouvertes en recherche.

La possibilité que les systèmes d'IA produisent des résultats inexacts constitue également un risque pour les applications en sciences et en génie. Les systèmes d'IA peuvent produire des résultats inexacts ou faussés à cause de biais dans les ensembles de données d'apprentissage et de problèmes de généralisation des données d'apprentissage aux nouvelles données. Plusieurs techniques permettent de détecter et d'éliminer les biais durant la phase d'apprentissage, ce qui peut éviter les conséquences et les résultats non désirés ou inexacts. Si certaines formes de biais sont inévitables, les développeurs peuvent être formés pour éviter les biais dans le recueil et la curation de données.

Même quand les résultats générés par l'IA sont exacts et non biaisés, plusieurs types populaires de systèmes d'IA fonctionnent comme des *boîtes noires*, qui rendent difficile, voire impossible, d'expliquer comment les résultats ont été produits. Cela peut entraver l'explication et la compréhension scientifiques et pourrait saper la crédibilité des constatations scientifiques produites par la machine. Bien que, dans certains cas, l'exactitude seule puisse être suffisante pour le progrès scientifique, l'objectif des sciences est ultimement l'explication et la compréhension, et il existe donc de nombreuses raisons de préférer des modèles d'IA interprétables aux boîtes noires dans la recherche en sciences et en génie.

Les facteurs éthiques à propos de l'utilisation de l'IA en sciences et en génie surgissent à tous les stades de la recherche, que ce soit au recueil et au prétraitement des données, à la conception et au déploiement des modèles d'IA entraînés avec ces données, à la diffusion et à la publication des résultats ou encore au stockage, à la tenue à jour et à l'accès à long terme aux données, modèles et résultats. L'IA aura donc une incidence sur les politiques institutionnelles touchant la conduite responsable et éthique de la recherche. De plus, la recherche sur l'IA complique la notion traditionnelle de consentement quand des participants humains sont concernés, car elle peut reposer sur des ensembles de données contenant des renseignements sur des personnes sans que celles-ci le sachent ou y aient consenti. Il sera nécessaire d'implanter des principes de gérance et de gestion des données pour faciliter le partage et l'utilisation responsables et éthiques de ces données. Il s'agit notamment des principes FAIR bien connus ainsi que d'autres principes de gestion des données complémentaires, tels que TRUST, FACT et, dans le cas de données concernant les peuples autochtones, CARE.

Le milieu de la R-D au Canada sera confronté à des défis dans l'adaptation de ses pratiques à l'évaluation de la recherche et des chercheurs utilisant l'IA

Un nombre croissant de pays, dont le Canada, ont élaboré des stratégies nationales d'amélioration de la capacité d'IA. La stratégie canadienne s'est traditionnellement concentrée sur la constitution d'une masse critique de chercheurs en IA, mais a commencé à cibler plusieurs secteurs et à tenir compte plus explicitement des



« L'utilisation de l'IA dans le processus de recherche brouille les limites entre les disciplines, ce qui rend les processus concurrentiels de financement, les examens scientifiques et les programmes de recherche axés sur un seul sujet moins utiles parce que l'expertise dans un domaine ne permet pas d'examiner correctement la nature interdisciplinaire de la science reposant sur l'IA. »

impacts sociétaux. L'élargissement de la portée des stratégies nationales d'IA peut être bénéfique parce que de nouvelles relations devront être établies dans le réseau de la R-D pour appliquer les connaissances spécialisées en IA aux problèmes des sciences et du génie. La recherche scientifique qui utilise l'IA peut également coïncider avec les objectifs et initiatives politiques existants, qui pourraient indirectement influencer la trajectoire du développement de cette technologie.

Le renforcement du rôle de l'IA dans la R-D en sciences et en génie ne se limitera pas aux laboratoires. Les effets seront également ressentis dans le milieu de la découverte et de l'innovation. L'utilisation de l'IA dans le processus de recherche brouille les limites entre les disciplines, ce qui rend les processus concurrentiels de financement, les examens scientifiques et les programmes de recherche axés sur un seul sujet moins utiles parce que l'expertise dans un domaine ne permet pas d'examiner correctement la nature interdisciplinaire de la science reposant sur l'IA. Par conséquent, les

sciences humaines et sciences sociales jouent un rôle important dans la R-D scientifique et en génie qui a recours à l'IA. Il se peut que l'on ait à repenser les cloisons disciplinaires traditionnelles du financement de la recherche pour garantir une évaluation équitable et appropriée de la recherche au moyen de l'IA.

Les promesses d'exactitude et de cohérence que recèle l'IA transforment le concept de responsabilité du chercheur. Si les systèmes d'IA deviennent la norme pour l'exécution des tâches scientifiques, les attentes concernant la responsabilité en matière d'exactitude et d'exhaustivité pourraient éliminer les travaux effectués par les humains de toute considération dans les concours de financement de la recherche.

Cela pourrait avoir comme effet de disqualifier les demandeurs qui n'ont pas accès à l'IA. L'emploi de l'IA dans la recherche complexifie des problèmes tels que la reproductibilité, l'explicabilité et l'exactitude, qui nécessiteraient une mise à jour du Cadre de référence des trois organismes sur la conduite responsable de la recherche.

La capacité de l'IA à prédire les répercussions de la recherche scientifique peut être utile pour les gouvernements et les organismes de financement qui décident s'il faut — et comment — offrir des possibilités de financement à la communauté scientifique. De plus, elle peut appuyer le processus d'examen par les pairs utilisé pour évaluer les applications scientifiques prospectives. Cependant, ces utilisations potentielles devront être soigneusement vérifiées afin de s'assurer qu'elles n'entraînent pas de conséquences indésirables. Par exemple, on s'inquiète du fait que l'IA puisse exacerber la marginalisation des groupes généralement sous-représentés dans le système de recherche. Bien que l'intégration de l'IA dans les systèmes de financement de la recherche et d'examen par les pairs puisse atténuer les problèmes de longue date concernant l'équité, la diversité et l'inclusion, on ne peut pas présumer que ce sera le cas. Une mise en œuvre minutieuse, qui tient compte de ces limitations, sera nécessaire.

Le déploiement accru des systèmes d'IA en sciences et en génie risque de perpétuer la discrimination ou les biais, que ce soit dans le milieu de la R-D canadienne ou dans l'ensemble de la société

Pour que l'IA soit utilisée de manière responsable en sciences et en génie, elle doit éviter de perpétuer les biais et la discrimination contre des individus ou des groupes; pourtant, il a déjà été observé que des outils d'IA amplifiaient ces problèmes dans de nombreux cas concrets. Le manque de diversité de genre et de race dans le domaine de la recherche en IA est bien documenté, et il existe actuellement une forte inégalité dans la répartition des ressources, de l'infrastructure et des compétences pour la production, la diffusion et l'utilisation de l'IA dans la recherche scientifique. Ce « fossé numérique » pourrait s'élargir en raison du fort coût des ressources de calcul et de l'intensification de la concurrence, surtout quand les investissements publics bénéficient principalement au secteur privé plutôt qu'aux universités ou au secteur public. Cette possibilité soulève la crainte de monopolisation du savoir scientifique. De plus, comme l'IA peut être utilisée pour la prise de décision en sciences et en génie — par exemple, pour les examens par les pairs ou les décisions de financement —, elle a de vastes implications sociales, comme la détermination des problèmes auxquels la recherche doit s'atteler.

L'utilisation de l'IA aura probablement de vastes répercussions sociales sur le marché du travail en sciences et en génie, sur la confiance du public dans l'IA et dans les sciences, sur l'environnement et sur la cybersécurité

Le manque de confiance dans l'IA pourrait entraver son adoption en sciences et en génie. Pour surmonter cet obstacle, les concepteurs de systèmes d'IA devront bâtir des données probantes fiables en démontrant de façon transparente des résultats fructueux et reproductibles. Le manque de confiance du public dans l'IA dans d'autres domaines pourrait nuire à la perception que l'IA pour les sciences et le génie est digne de confiance, tout comme les résultats biaisés ou discriminatoires ou les pratiques non éthiques concernant le recueil et l'utilisation des données. La confiance dans l'IA en sciences et en génie exigera aussi de s'occuper de la sécurité des systèmes d'IA et des risques pour les propriétaires, les utilisateurs et les personnes touchées par ces systèmes.

L'augmentation de l'utilisation de l'IA pourrait également avoir une incidence sur le marché du travail en sciences et en génie. Certaines suppressions d'emplois sont inévitables; cependant, le principal effet de l'IA sur l'emploi dans ces secteurs sera la transformation des postes. L'acquisition de compétences et la formation seront nécessaires pour adapter la main-d'œuvre au changement produit par le recours accru à l'IA. Cela nécessitera de développer les connaissances et les compétences techniques nécessaires pour innover et déployer l'IA à des fins de recherche, et de fournir aux futurs scientifiques et ingénieurs la capacité de déterminer et de résoudre les facteurs sociaux et éthiques associés au développement et à l'utilisation de l'IA.

Plus généralement, les tiraillements continuent à croître autour des impacts environnementaux de l'IA. D'un côté, les découvertes effectuées à l'aide de cette technologie pourraient aider à lutter contre la crise climatique. De l'autre, le développement et le fonctionnement des systèmes d'IA demandent une quantité considérable d'énergie et peuvent produire énormément d'émissions de gaz à effet de serre. Et au chapitre des impacts environnementaux de l'IA, il faut aussi inclure ceux découlant de l'extraction des matières premières et de la fabrication de composants, le transport des matériaux, la construction et l'installation de l'infrastructure, l'entretien, la réparation, la remise à neuf et la mise à niveau du système et les opérations en fin de vie, comme le transport, le traitement des déchets et l'élimination.

Le développement technologique s'effectue à une cadence plus élevée que l'élaboration des cadres légaux et réglementaires qui régissent les systèmes d'IA, ce qui conduit à une incertitude concernant le déploiement et la commercialisation

Dans le domaine légal, plusieurs barrières émergent autour de l'utilisation de l'IA en sciences et en génie, surtout dans le champ de la propriété intellectuelle. Les cadres de propriété intellectuelle traditionnels sont difficiles à appliquer aux machines capables d'innovation ou d'originalité. Les brevets et les droits d'auteur sont devenus des actifs essentiels dans une économie de plus en plus numérique et incorporelle; or, ces instruments étaient à l'origine conçus comme des encouragements destinés aux humains. Ce problème intéresse particulièrement le Canada, étant donné les récentes initiatives visant à accroître le rendement de l'innovation.

La propriété ou le contrôle des données — ainsi que l'accès aux données — émerge comme des préoccupations essentielles dans le développement de l'IA. Les politiques facilitant l'accès aux données et aux données ouvertes et promouvant la gouvernance et la gestion responsables des données sont importantes pour l'IA en sciences et en génie; cependant, des lacunes réglementaires et la fragmentation résultant de la complexité du cadre légal régissant les données au Canada peuvent entraver l'accès à certains types de renseignements. Une contradiction existe également entre la transparence et la confidentialité, en particulier du point de vue de la commercialisation. Les innovateurs peuvent être prudents lorsque vient le temps de partager des données précieuses, car ils peuvent être mieux protégés par la confidentialité que par d'autres mécanismes formels de protection de la propriété intellectuelle. Certains de ces dilemmes peuvent être résolus par des initiatives de données ouvertes et de gouvernance des données, combinées à des réformes législatives, bien que leur conception ne soit pas nécessairement coordonnée.

Au-delà de la propriété intellectuelle, la responsabilité légale demeure un autre secteur qui fait débat. Les cadres actuels ont du mal avec l'attribution de la responsabilité des préjudices découlant de l'utilisation d'un système d'IA à des fins prédictives ou décisionnelles, ce qui augmente les risques dans les domaines de la confiance et de la reddition de comptes. Les gouvernements et les décideurs devront déterminer s'il faut adapter les cadres actuels de responsabilité aux systèmes d'IA ou s'ils doivent élaborer de nouveaux cadres qui s'attaquent aux incertitudes entourant l'IA.

Il n'existe pas de loi unique pour la réglementation ou la gouvernance de l'IA au Canada; la division des pouvoirs entre le fédéral et les provinces ou les territoires présente des défis pour la création d'un cadre réglementaire unique. Les lois régissant les données personnelles à l'échelon provincial ou territorial et à l'échelon fédéral ont commencé à porter leur attention sur l'utilisation des données personnelles dans la prise de décision automatisée et pourraient s'orienter sur les données anonymisées. De nombreuses activités en sciences et en génie ne sont donc pas actuellement réglementées parce qu'elles pourraient ne pas inclure de données personnelles. La fragmentation qui en résulte constitue un problème en raison de la nature collaborative du développement de l'IA, ce qui a des conséquences sur la commercialisation dans le cadre de partenariats public-privé et de partenariats internationaux.

Malgré les efforts en cours pour préparer le Canada afin qu'il soit compétitif sur le plan économique dans un environnement réglementaire changeant, le développement et l'application responsables et éthiques de l'IA continuent à se heurter à des obstacles substantiels, en raison de l'incertitude concernant l'opérationnalisation et la prise en considération de la technologie. Aborder les systèmes d'IA comme des systèmes sociotechniques pourrait permettre de surmonter ces défis en élargissant la portée de l'étude afin d'inclure des développements qui semblent périphériques aux applications d'IA en question.

L'utilisation de systèmes d'IA en sciences et en génie pousse les limites disciplinaires, la collaboration et la coordination vers un futur transdisciplinaire

Bien que certains défis légaux ou réglementaires, éthiques, sociaux et politiques découlant de l'utilisation de l'IA en sciences et en génie puissent être réglés par la reconfiguration des systèmes techniques eux-mêmes, ce rapport suggère que l'accroissement des travaux transdisciplinaires sera nécessaire. La vision étroite des sciences et du génie — selon laquelle les disciplines des sciences, de la technologie, du génie et des mathématiques demeurent distinctes des sciences sociales, des sciences humaines et des sciences de la santé — est de plus en plus inappropriée pour améliorer la compréhension et l'explication.

En tant qu'outil de recherche de plus en plus utilisé par de plus en plus de gens à de plus en plus de stades du processus de découverte, l'IA transforme les bases épistémiques des sciences et du génie. Les méthodes et les procédures qui font en sorte que le savoir produit au moyen de l'IA est digne de confiance — et exact,



« La vision étroite des sciences et du génie — selon laquelle les disciplines des sciences, de la technologie, du génie et des mathématiques demeurent distinctes des sciences sociales, des sciences humaines et des sciences de la santé — est de plus en plus inappropriée pour améliorer la compréhension et l'explication. »

explicable et reproductible — auront besoin d'être révisées. Il faudra donc répondre aux préoccupations relatives à la transparence des systèmes d'IA. De plus, l'intégrité du système de recherche canadien dépendra également du déploiement responsable des systèmes d'IA, et harmonisé avec les attentes évolutives concernant l'éthique, l'équité, la diversité et l'inclusion.

Quand les avancées rendues possibles par les systèmes d'IA utilisés en sciences et en génie sont mises en pratique, elles peuvent améliorer la vie de certains, mais également causer des préjudices à d'autres. Les cadres légaux et politiques devront être modifiés avec l'IA à l'esprit pour concilier la sécurité et le bien-être avec l'innovation.

Toutefois, on ne sait pas encore comment résoudre les problèmes de propriété intellectuelle concernant les données et la découverte automatisée. Si les contrats sont actuellement utilisés pour établir les droits, les responsabilités

et autres entre les parties, une réforme réglementaire sera nécessaire pour que les scientifiques et les ingénieurs qui utilisent des systèmes d'IA bénéficient d'un certain degré de certitude.

Ces défis exigent une pensée et une collaboration transdisciplinaires. La stratégie nationale d'IA commence à appliquer ces concepts pour transformer l'écosystème de la recherche canadien en coordonnant les investissements, la réforme des politiques et la formation. L'approche transdisciplinaire devra également appuyer les changements nécessaires aux politiques de financement de la recherche et de formation si le Canada veut continuer à promouvoir l'éthique, l'équité, la diversité et l'inclusion en sciences et en génie.

Abréviations

La liste qui suit ne recense pas l'intégralité des abréviations utilisées dans ce report, mais présente les termes les plus couramment utilisés.

Amii	Alberta Machine Intelligence Institute
CIFAR	Institut canadien de recherches avancées
CNRC	Conseil national de recherches Canada
CRSH	Conseil de recherche en sciences humaines
CRSNG	Conseil de recherches en sciences naturelles et en génie du Canada
EDI	équité, diversité et inclusion
ETD	exploration de texte et de données
GES	gaz à effet de serre
GDR	gestion des données de recherche
IA	intelligence artificielle
IRSC	Instituts de recherche en santé du Canada
LESP	légal/réglementaire, éthique, social et politique
Mila	Quebec Artificial Intelligence Institute (anciennement le Montreal Institute for Learning Algorithms)
PDA	prise de décision automatisée
PME	petites et moyennes entreprises
R-D	recherche-développement
RDA	Research Data Alliance
XAI	intelligence artificielle explicable

Glossaire

Algorithme : « Formule ou ensemble de règles (ou procédure, processus ou instructions) permettant de résoudre un problème ou de réaliser une tâche. En intelligence artificielle, l'algorithme indique à l'ordinateur comment trouver des réponses à une question ou des solutions à un problème. En apprentissage machine, les systèmes utilisent différents types d'algorithme. Parmi les exemples courants, citons les arbres de décision, les algorithmes de groupement, les algorithmes de classification et les algorithmes de régression » [traduction libre] (Guo *et al.*, 2019).

Intelligence artificielle (IA) : Pour les besoins de ce rapport et pour permettre une interprétation plus large et inclusive des technologies d'intelligence artificielle, l'IA est définie comme un ensemble de techniques statistiques et logicielles, ainsi que les données associées et leur contexte social. De plus, le terme « intelligence artificielle » est utilisé de façon interchangeable pour décrire diverses implémentations (méthodes) de conception et de découverte assistées par l'ordinateur :

Apprentissage profond : « Sous-domaine de l'apprentissage machine qui se préoccupe des algorithmes inspirés par le cerveau humain et travaillant de façon hiérarchique. Les modèles d'apprentissage profond, qui s'appuient principalement sur les réseaux neuronaux (artificiels), sont appliqués à différents domaines, tels que la reconnaissance de la parole, la vision par ordinateur et le traitement automatique des langues » [traduction libre] (Guo *et al.*, 2019).

Apprentissage machine : « Renvoie normalement à la branche de l'IA portant sur la conception de systèmes qui apprennent à partir des données. Au lieu qu'on leur dise explicitement comment résoudre un problème, les algorithmes [d'apprentissage machine] peuvent créer des solutions en apprenant à partir d'exemples (ce qu'on appelle "apprentissage" des algorithmes [d'apprentissage machine]) » [traduction libre] (King et Roberts, 2018).

Apprentissage par renforcement : « Type de programmation dynamique qui entraîne les algorithmes au moyen d'un système de récompenses et de punitions. L'algorithme est exposé à un ensemble de données totalement [sic] aléatoire et nouveau et détermine automatiquement des modèles et des relations à l'intérieur de cet ensemble de données. Le système est récompensé quand il trouve une relation souhaitée au sein de l'ensemble, mais il est

aussi puni s'il trouve une relation non désirée. L'algorithme apprend des récompenses et des punitions et s'actualise lui-même continuellement. Ce type d'algorithme est toujours en mode production. Pour s'actualiser et présenter des actions, il a besoin de données en temps réel. L'agent apprend sans intervention de l'humain en maximisant les récompenses et en minimisant les punitions » [traduction libre] (Guo *et al.*, 2019).

Chercheur autonome (scientifique autonome) : « Regroupement d'une multitude de modules logiciels et matériels qui interagissent dynamiquement pour accomplir des tâches [et] qui sont capables d'effectuer des recherches autonomes pour réaliser des découvertes scientifiques majeures » [traduction libre] (Kitano, 2021).

Biais :

Biais d'automatisation : Se produit quand les gens font confiance aux « systèmes de soutien automatisés. Cette confiance est due à la perception chez les humains que ces systèmes ont des capacités analytiques supérieures à celles de leurs homologues humains » [traduction libre] (Lopez *et al.*, 2019). En revanche, on parle d'**aversion aux algorithmes** quand les chercheurs choisissent des prédictions humaines au détriment de celles produites par les algorithmes, même après avoir eu la preuve que les prédictions humaines sont moins exactes (Dietvorst *et al.*, 2015).

Ensembles de données ou modèles biaisés : Quand un système d'IA produit des résultats faussés ou inexacts (y compris des résultats discriminatoires) pour diverses raisons, comme des biais dans les ensembles de données dus à une discrimination historique préexistante ou à des erreurs d'échantillonnage ou de prétraitement des données, ou à des décisions subjectives prises par les chercheurs quand ils mettent au point le modèle, comme le choix inadéquat du modèle ou des paramètres du modèle, l'absence de supervision humaine et le manque de transparence (Veale et Binns, 2017; WEF, 2018).

Biais de discrimination : Dans l'usage courant et dans de nombreux comptes rendus médiatiques, le terme biais fait généralement référence aux modèles d'IA qui produisent des résultats discriminatoires (Hellström *et al.*, 2020). C'est-à-dire, quand « des jugements injustes sont posés parce que la personne qui les pose est influencée par une caractéristique qui n'est *en fait* pas rattachée au sujet, généralement une préconception discriminatoire des membres d'un groupe » [traduction libre] (Muller, 2020).

Parti pris inductif : Un des nombreux types de biais nécessaires pour que l'IA fonctionne. Le parti pris inductif permet à l'IA de généraliser à de nouveaux exemples à partir de ses données d'apprentissage parce que « l'apprentissage efficace à partir de données finies exige des hypothèses quant à la source de données [...], car quelle que soit la quantité finie de données provenant d'une source inconnue, il est impossible de prédire l'élément suivant, même approximativement, à moins de disposer de certains renseignements mineurs à propos de cette source » [traduction libre] (Amit et Meir, 2019).

Boîte noire : Système d'IA dont les intrants et les extrants sont connus, mais dont le fonctionnement interne n'est pas compris. « Une fois entraîné, [un système d'IA boîte noire] peut produire des résultats statistiquement fiables, mais l'utilisateur final ne sera pas nécessairement en mesure d'expliquer comment ces résultats ont été produits ou de déterminer les caractéristiques particulières d'un cas qui ont joué un rôle important dans l'atteinte de la décision finale » [traduction libre] (The Royal Society, 2019).

Données :

Mégadonnées : « Vastes ensembles de données produits sous forme numérique et qui peuvent être analysés au moyen d'outils de calcul » [traduction libre] (Leonelli, 2020).

Provenance des données : Permet « d'examiner l'historique et le processus de construction des ensembles de données et d'étudier comment les normes et les stéréotypes culturels ont été énumérés et représentés au moment de la création de ces données » [traduction libre] (West *et al.*, 2019). La provenance des données est souvent utilisée pour atténuer ou éviter les biais possibles dans les ensembles de données.

Metadonnées : Données qui renseignent sur d'autres données.

Données ouvertes : « [D]onnées structurées lisibles par machine, librement partagées, utilisées et exploitées sans restriction » (GC, 2020a).

Éthique de l'IA : « Sous-domaine de l'éthique appliquée qui se concentre sur les questions d'éthique soulevées par le développement et l'utilisation de l'IA. Elle se préoccupe principalement de déterminer comment l'IA peut améliorer le bien-être des individus ou soulever des craintes à son sujet, que ce soit sur le plan de la qualité de vie ou de l'autonomie et de la liberté humaines nécessaires à une société démocratique » [traduction libre] (IA HLEG, 2019).

IA explicable (XAI) : « Efforts visant à s’assurer que les programmes d’intelligence artificielle sont transparents dans leurs objectifs et dans leur façon de fonctionner. L’IA explicable est un objectif commun aux ingénieurs et autres qui essaient de faire progresser l’intelligence artificielle » [traduction libre] (Guo *et al.*, 2019).

Souveraineté des données autochtones : « Faculté des peuples, des communautés et des Nations autochtones à contribuer aux données qu’ils créent ou qui les concernent, et à les gérer et les contrôler. Le terme *souveraineté* fait référence au fait que les Nations autochtones sont souveraines dans leur gouvernance, ainsi que relativement à leurs données et à leur savoir. Il reconnaît que les peuples autochtones sont l’autorité ultime en ce qui concerne leurs données et leur savoir et vise à redéfinir la relation de ces peuples avec la recherche pour que de simples participants ou sujets, ils deviennent des partenaires importants et des cochercheurs » [traduction libre] (UofT, 2021a).

Interprétabilité (ou explicabilité) : Peut être généralement interprété comme l’aptitude des humains à comprendre comment un modèle d’IA donné fonctionne et pourquoi il produit les résultats qu’il produit (Lipton, 2018; Rudin, 2019). Les modèles d’IA interprétables ou explicables (ces termes sont utilisés de façon interchangeable dans le rapport) contrastent avec les modèles boîte noire.

Science ouverte : « [C]oncept inclusif qui englobe différents mouvements et pratiques visant à rendre les connaissances scientifiques multilingues librement accessibles à tous et réutilisables par tous, à renforcer la collaboration scientifique et le partage des informations au profit de la science et de la société, ainsi qu’à ouvrir les processus de création, d’évaluation et de diffusion des connaissances scientifiques aux acteurs de la société au-delà de la communauté scientifique traditionnelle » (UNESCO, 2020).

Reproductibilité (ou répliquabilité) : Degré auquel la répétition d’une expérience donne des résultats cohérents. Un des principaux moyens pour la communauté scientifique de valider l’exacitude de découvertes ou de nouvelles constatations, elle est considérée comme une des « caractéristiques d’une bonne démarche scientifique » [traduction libre] (NASEM, 2019). Il est cependant à noter que des termes tels que *reproductibilité* et *répliquabilité* sont parfois définis et utilisés de façon distincte et même contradictoire dans différentes disciplines (Fiddler et Wilcox, 2018; NASEM, 2019). En fait, même dans le domaine de la recherche en IA, ces deux termes peuvent avoir des significations différentes selon la source. Toutefois, dans ce rapport, *reproductibilité* est simplement utilisé comme un terme général recouvrant une grande variété de situations, de manière interchangeable avec *répliquabilité*.

Table des matières

1	Introduction	1
1.1	Mandat du comité d'experts	4
1.2	Approche du comité d'experts	4
1.3	Portée du rapport	6
1.4	Structure du rapport	7
2	L'IA dans le contexte des sciences et du génie	8
2.1	Exploitation de l'IA pour la conception et la découverte . .	10
2.2	Contexte canadien	17
2.3	Progrès en gouvernance et en politique	20
3	Répercussions de l'IA sur la conduite de la recherche . .	24
3.1	Garantie de l'intégrité scientifique dans la recherche utilisant l'IA	26
3.2	Utilisation éthique de l'IA dans la recherche en sciences et en génie	37
3.3	Les répercussions de l'IA sur les valeurs et sur les pratiques sociales en sciences	45
4	Conséquences de l'IA sur le système de recherche canadien	48
4.1	L'utilisation de l'IA dans l'allocation de ressources recherche	50
4.2	Adaptation des systèmes d'examen par les pairs à l'IA . . .	53
4.3	Mesure de l'impact de la recherche	56
4.4	Intégrité de la recherche et gouvernance de la conduite de la recherche	59
4.5	Formation et acquisition des compétences	61

5	Les conséquences sociales de la recherche employant l'IA	64
5.1	Confiance dans l'IA : biais d'automatisation et aversion aux algorithmes	66
5.2	Gérer le biais et la discrimination dans les systèmes d'IA	68
5.3	Impact de l'IA sur les communautés autochtones.	71
5.4	L'EDI dans la recherche sur l'IA et dans l'accès à la technologie d'IA	74
5.5	Répercussion sur le marché du travail	77
5.6	Impacts environnementaux des systèmes d'IA	79
5.7	Sécurité de l'IA	81
6	Conséquences de l'IA pour les lois, la réglementation et la politique canadiennes	84
6.1	Accès aux données	86
6.2	Commercialisation des résultats produits par la machine	93
6.3	Nouveaux risques légaux	104
6.4	Nouveaux systèmes réglementaires pour l'IA.	107
7	Conclusion.	114
7.1	Réponse au mandat.	115
7.2	Réflexions du comité d'experts.	120
	Références	123

Introduction

- 1.1 Mandat du comité d'experts
- 1.2 Approche du comité d'experts
- 1.3 Portée du rapport
- 1.4 Structure du rapport

Bien que le potentiel de l'intelligence artificielle (IA) de perturber la société continue à frapper l'imagination du public et des décideurs, les progrès de l'IA pour ce qui est de soutenir la conception et les découvertes dans la recherche en sciences et en génie attirent moins l'attention. Ces avancées sont néanmoins rapides, tout comme l'évolution de leurs répercussions sur ces domaines.

L'IA commence déjà à stimuler la recherche scientifique en permettant la génération d'hypothèses, la conception d'expériences, l'expérimentation, l'interprétation et l'analyse automatiques. Les progrès techniques et la disponibilité accrue des données ont favorisé le développement et le déploiement de systèmes d'IA pour nombre de ces processus. Des effets similaires pourraient bientôt toucher le génie, où l'amélioration continue de l'IA conjuguée aux avancées



« En fin de compte, la question n'est pas de savoir si l'IA finira par jouer un rôle égal ou supérieur à celui de l'humain dans la conception et la découverte en sciences et en génie, mais avec quelle rapidité. »

en robotique et dans d'autres technologies reliées pourrait radicalement transformer la conception et la fabrication industrielles. Les outils d'IA sont déjà largement utilisés dans certains milieux et disciplines spécialisés et les percées en IA et dans son application dans des domaines distincts continuent à se multiplier. En fin de compte, la question n'est pas de savoir si l'IA finira par jouer un rôle égal ou supérieur à celui de l'humain dans la conception et la découverte en sciences et en génie, mais *avec quelle rapidité*.

Les applications et les utilisations de l'IA sont diverses et suivent de nombreuses typologies. Les systèmes d'IA modernes sont composés de matériel contrôlé par un logiciel utilisant les données, mais

ces systèmes sont écrits, assemblés et maintenus par des humains, puis déployés pour remplir une fonction déterminée par des » humains. L'IA est une plateforme technologique composée de systèmes sociotechniques reliés à leur environnement de diverses façons, qui s'étendent de la formation des développeurs humains aux données que ces développeurs utilisent, en passant par la façon dont ces données

sont recueillies et celle dont chaque système évolue en fonction de son environnement, voire l'inverse. Les utilisations de l'IA ont aussi tendance à varier selon les applications, les disciplines et les sources de données d'intérêt (The Royal Society, 2019). En réalité, même si les praticiens de diverses disciplines concevront des outils d'IA correspondant à leurs propres besoins, les progrès réalisés dans chaque domaine distinct serviront collectivement à faire avancer la science de l'IA dans son ensemble (The Royal Society, 2019), mais cela pourrait compliquer les efforts de prédiction de l'incidence générale des avancées individuelles réalisées dans des silos disciplinaires.

Un cadre légal, politique et réglementaire de l'IA — quoique fragmenté — commence à émerger, bien que certaines interrogations demeurent. L'accélération mondiale de l'activité en matière d'IA a mis en lumière plusieurs défis éthiques liés à des biais et des pratiques préjudiciables. Bien que l'on continue à prévoir et à encourager le développement constant et les possibilités offertes par l'IA, il sera vital, à mesure que cette technologie est introduite dans la société, de s'assurer que son développement s'accompagne d'une réflexion sur son impact sociétal. Cette réflexion doit s'effectuer durant les phases de conception et de développement technologiques, plutôt qu'après la mise en œuvre.

Reconnaissant les possibilités, les défis et les conséquences de l'IA en sciences et en génie, le Conseil national de recherches Canada (CNRC), conjointement avec l'Institut canadien de recherches avancées (CIFAR), les Instituts de recherche en santé du Canada (IRSC), le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et le Conseil de recherche en sciences humaines (CRSH), a demandé au Conseil des académies canadiennes (CAC) de former un comité d'experts pour effectuer une évaluation fondée sur les données probantes et faisant autorité des conséquences légales ou réglementaires, éthiques, sociales et politiques de l'utilisation de l'IA dans ces deux domaines.

1.1 Mandat du comité d'experts

Le CAC avait pour mandat de répondre à la question et aux sous-questions suivantes :



Quels sont les défis légaux/réglementaires, éthiques, sociaux et politiques du déploiement des technologies d'IA pour permettre la conception de la recherche et la découverte en sciences et en génie au Canada?

- Comment l'utilisation de l'IA dans la recherche en sciences et en génie modifie-t-elle ou influence-t-elle les pratiques, processus et résultats standards dans ces deux domaines? Notamment :
 - Comment assurer l'intégrité scientifique (p. ex. la reproductibilité et la validité)?
 - Quelles sont les répercussions sociétales, éthiques et épistémiques dont il faut tenir compte?
- Quelles sont les dimensions sociales et éthiques que les chercheurs individuels et les collaborations de recherche doivent prendre en compte lorsqu'ils utilisent l'IA en sciences et en génie (p. ex. les biais, le sexe ou le genre)? Comment soutenir les générations actuelles et futures de chercheurs pour garantir des pratiques éthiques?
- Quelles sont les politiques concernant le déploiement et l'utilisation de l'IA en sciences et en génie mises en œuvre dans les provinces et territoires canadiens et à l'étranger?

1.2 Approche du comité d'experts

Pour répondre à ce mandat, le CAC a mis sur pied un comité multidisciplinaire et multisectoriel d'experts provenant du Canada et de l'étranger. Les travaux du comité se sont déroulés pendant la pandémie de COVID-19 et ont donc dû s'effectuer de manière virtuelle. Le comité d'experts s'est réuni virtuellement cinq fois en huit mois afin d'examiner les données probantes et de délibérer sur son mandat. Comme pour toutes les évaluations du CAC, son rapport a été examiné par des pairs avant sa publication. Il s'est focalisé sur les défis légaux ou réglementaires, éthiques, sociaux et politiques (LESP), et notamment sur l'intégrité scientifique et sur les questions épistémiques et éthiques liées au déploiement des technologies d'IA pour permettre la conception de la recherche et la découverte en sciences et en génie au Canada.

1.2.1 Sources des données probantes

Le comité d'experts admet qu'il existe de nombreux articles, revues et rapports portant sur l'identification des problèmes de LESP concernant l'IA. Bien que le domaine sur lequel ce rapport se concentre — l'IA en sciences et en génie — ait été moins étudié, il a décidé, quand c'était possible, d'examiner l'opérationnalisation des approches permettant de s'attaquer aux problèmes précédemment relevés. Le comité souhaite à ce titre attirer l'attention sur *Learning: The Power and Promise of Computers That Learn by Example*, une publication fondatrice de la Société royale et du Turing Institute, du Royaume-Uni, parue en avril 2017.

Le comité d'experts a effectué un examen semi-structuré des publications par le biais de la base de données de la bibliothèque Web of Science, au moyen d'un logiciel bibliographique en source ouverte (VOSViewer) pour déterminer itérativement les mots-clés relatifs aux récentes publications examinées par les pairs intéressantes pour le mandat. La majorité des publications tirées de l'examen dépassaient le cadre des sciences et du génie et portaient sur les aspects techniques des systèmes d'IA ou sur des facettes de la LESP touchant les applications de cette technologie et concernant la société de façon plus générale. Par conséquent, la documentation parallèle, qui se compose de documents de politique, de publications gouvernementales et de rapports publiés par des organismes sans but lucratif (OSBL) nationaux et internationaux, a constitué une source importante de données probantes pour la rédaction de ce rapport.

1.2.2 Public cible

Les questions étudiées dans ce rapport touchent un vaste éventail d'acteurs au Canada et à l'étranger. Le comité d'experts prévoit que ce rapport intéressera les personnes responsables de l'élaboration des politiques, les décideurs des secteurs de l'IA, les organismes gouvernementaux engagés dans l'IA et dans ses applications, les OSBL et les chercheurs industriels et universitaires, en ce qui a trait aux risques, aux avantages, aux défis et aux facteurs qui accompagnent le développement et le déploiement de l'IA en sciences et en génie.

1.2.3 Interprétation de l'intelligence artificielle par le comité d'experts

La définition de l'IA est fluide et reflète un paysage en constante évolution marqué par les avancées technologiques et l'élargissement des domaines d'application. En fait, on a fréquemment observé qu'une fois que l'IA est en mesure de résoudre un problème donné ou d'accomplir une certaine tâche, elle n'est souvent plus considérée comme de l'intelligence « véritable » (Haenlein et Kaplan, 2019). En raison de la variété des implantations décrite ci-dessus, le comité n'a pas appliqué de définition stricte pour ce rapport. Cependant, pour les besoins de la délibération, il a choisi de considérer l'IA comme une collection de techniques statistiques et logicielles, ainsi que leurs données associées et leur contexte social, ce qui permet une interprétation plus vaste et inclusive des technologies et des formes de capacité d'agir. Le comité d'experts utilise le terme IA de façon indifférenciée pour décrire diverses implémentations de conception et de découverte assistée par la machine, notamment celles fondées sur l'apprentissage machine, l'apprentissage profond et l'apprentissage par renforcement, à l'exception d'exemples particuliers dans lesquels le choix de la mise en œuvre est marquant.

1.3 Portée du rapport

Les premières discussions avec les commanditaires ont clarifié la portée et les objectifs du rapport. Pour ceux-ci, il était important de savoir comment l'IA, appliquée à la conception et à la découverte, pouvait mener à de nouvelles démarches et de nouveaux paradigmes en sciences et en génie, qui complèteront et, potentiellement, remplaceront les pratiques standards conventionnelles. Bien que les commanditaires aient insisté sur les sciences et le génie dans leur mandat, celui-ci recouvre aussi résolument la recherche en sciences sociales, dans la mesure où les éléments de LESP sont communs et inhérents à toutes les disciplines de ce domaine.

La portée de ce rapport inclut la recherche et le développement (R-D) en création et en utilisation des outils d'IA pour la conception et la découverte dans les secteurs universitaire, privé et gouvernemental. La recherche axée sur la découverte est incluse dans la portée du rapport, tout comme la R-D commerciale, en raison du rôle actif joué par le secteur privé dans la recherche, le développement et le déploiement de l'IA. C'est le cas également de la recherche biomédicale élémentaire, en raison des promesses affichées par les techniques d'apprentissage machine matière de navigation dans le complexe espace d'élaboration des questions de recherche intéressant le domaine biomédical.

Toutefois, le rapport n'offre pas de compte rendu intégral des questions de LESP concernant l'IA appliquée à la sécurité ou aux soins de santé. D'autres travaux réalisés ou en cours de réalisation offrent un aperçu de l'utilisation de l'IA en recherche sur la santé incluant la participation directe de sujets humains; par conséquent, le comité d'experts a cherché à éviter ce domaine autant que possible afin de ne pas étendre démesurément son mandat. Suivant la terminologie employée par les IRSC, ce rapport accorde la priorité aux questions concernant le premier thème de la recherche (recherche biomédicale), mais étudie également des domaines qui chevauchent les trois autres thèmes, le cas échéant¹.

1.4 Structure du rapport

Le **chapitre 2** décrit le contexte — et la motivation — de l'utilisation de l'IA dans la conception et la découverte en sciences et en génie. Il dresse également un aperçu de l'écosystème canadien d'intelligence artificielle, ainsi que de l'évolution internationale pertinente concernant la croissance de la gouvernance de l'IA. Le **chapitre 3** présente les concepts théoriques et fondamentaux cruciaux pour le déploiement de l'IA en sciences et en génie et les répercussions qu'elle pourrait avoir sur la nature de l'enquête scientifique et de la conception en génie, tels que la reproductibilité, l'interprétabilité, le biais, le comportement éthique en recherche et la pratique sociale des sciences. Le **chapitre 4** examine les conséquences de l'utilisation des technologies d'IA sur les processus de financement de la recherche en sciences et en génie, ainsi que les effets possibles sur les politiques qui régissent la recherche et l'éducation.

Les **chapitres 5 et 6** portent sur les défis et les possibilités du déploiement des technologies d'IA qui permettent la conception et la découverte en sciences et en génie. Le **chapitre 5** recense les implications sociales et politiques, notamment sur la confiance de la population, la sécurité de l'apprentissage machine, les répercussions sur le marché du travail et sur l'environnement et la gestion des biais dans la classification des données. Le **chapitre 6** répertorie les problèmes légaux et réglementaires, notamment en matière d'accès aux données, de commercialisation de l'innovation, de responsabilité et de sécurité. Le **chapitre 7** conclut ce rapport en proposant un résumé des constatations du comité d'experts et ses réflexions finales au sujet de son mandat.

1 Les quatre thèmes des IRSC en ce qui concerne la recherche en santé sont la recherche biomédicale, la recherche clinique, les systèmes et les services de santé et la santé des populations (IRSC, 2021).

L'IA dans le contexte des sciences et du génie

- 2.1 Exploitation de l'IA pour la conception et la découverte
- 2.2 Le contexte canadien
- 2.3 Progrès en gouvernance et en politique

Constatations du chapitre

- Les progrès en IA et dans les technologies connexes multiplieront les possibilités d'automatisation en sciences et en génie, préparant ainsi le terrain pour l'innovation autonome et présentant un vaste éventail de conséquences pour la R-D.
- La stratégie du Canada en IA a abouti à la création de plusieurs pôles de recherche bien développés, qui jettent les bases d'occasions de développement économique lié à l'IA.
- L'IA émerge comme une plateforme technologique et nombre d'intervenants, de politiques et de tendances extérieurs à son environnement influenceront les activités qui y sont menées. Il faudra donc être conscient des évolutions dans les secteurs voisins pour pouvoir mener efficacement les activités en sciences et en génie.
- Malgré la rapide évolution du contexte réglementaire international, les efforts visant à établir des directives robustes pour le développement responsable et éthique de l'IA sont entravés par l'incertitude en matière de gouvernance et de reddition de comptes.

L'IA en sciences et en génie vit une croissance et des changements rapides. Portée par un volume grandissant de données, des ressources informatiques bon marché et un élargissement de l'éventail d'applications à tous les stades de la conception et de la découverte, elle est désormais un outil très répandu en sciences et en génie. Grâce à la combinaison d'algorithmes et de sources de données, les systèmes d'IA sont portables et incorporels comparés aux technologies et outils transformateurs précédents, dont l'emplacement et la forme pouvaient être fixes. Toutefois, réaliser l'équilibre entre encourager le développement de cette technologie — y compris les avantages sociaux et économiques connexes — et protéger les droits humains et s'assurer de l'utilisation éthique et responsable de l'IA tout au long de son cycle de vie se heurte à une divergence fondamentale. En fait, les décideurs du Canada doivent également être conscients des progrès réalisés à l'étranger parce que la mise sur pied de cadres réglementaires internationaux peut avoir des conséquences sur les avancées au pays, ce qui peut pousser les responsables des politiques à harmoniser les réglementations afin d'éviter la perturbation des activités commerciales internationales.

2.1 Exploitation de l'IA pour la conception et la découverte

Le but initial de l'IA était d'acquérir l'information de la façon dont le fait le cerveau et de la manipuler selon des règles codées dans des algorithmes pour parvenir à des décisions et des résultats avec une rapidité et une exactitude supérieures aux humains (Dick, 2019). Mais les méthodes modernes ont dépassé l'usage de systèmes spécialisés dans le stockage et le traitement de vastes quantités d'information sur certains sujets pour imiter la façon dont les humains prennent des décisions (Anyaha, 2017). L'IA, principalement par l'intermédiaire de systèmes d'apprentissage machine, englobe aujourd'hui un ensemble diversifié de techniques adaptatives et dynamiques, dont beaucoup se concentrent sur la création de prédictions exactes à partir des données (Dick, 2019), et a des effets sur de nombreuses formes de travail cognitif humain, y compris la recherche et l'application du savoir scientifique.

Les techniques d'IA actuellement utilisées en sciences et en génie appuient généralement la conception et la découverte effectuées par les humains

L'IA est employée dans la R-D élémentaire, appliquée et expérimentale depuis plusieurs années, particulièrement en analyse des données (The Turing Institute, 2021). Nombre des problèmes canoniques susceptibles d'être résolus au moyen de l'apprentissage machine — comme la classification, la régression et le groupement — sont largement applicables à la recherche scientifique (The Royal Society, 2017). La classification peut être largement utilisée dans l'analyse des images, par exemple, alors que les algorithmes de régression peuvent faciliter la création de modèles prédictifs reposant sur les données (The Royal Society, 2017). Les données numériques sont créées, recueillies, stockées et manipulées directement ou encore indirectement par l'intermédiaire d'outils informatiques ou de matériel contrôlé par ordinateur. Le volume et la diversité des données qui en résultent en sciences et en génie offrent de nombreux points d'insertion dans la boucle de conception et de découverte, dans laquelle les systèmes d'IA permettent ou accélèrent les progrès au-delà de l'analyse des données (Stevens *et al.*, 2020). Ces systèmes peuvent être davantage que des assistants ou des outils; ils pourraient être déployés pour poursuivre de façon autonome la découverte ascendante exploratoire, sans être limités par les biais cognitifs ou être guidés par les valeurs ou les contraintes sociologiques qui influent sur les chercheurs (Kitano, 2021).

Les utilisations actuelles de l'IA en sciences et en génie sont diverses et varient selon les applications, les disciplines et les sources de données d'intérêt (The Royal Society, 2019). Les outils d'IA peuvent être génératifs et utilisés pour produire des conjectures mathématiques abstraites (Castelvecchi, 2021), pour

concevoir des formes de vie artificielles (Kriegman *et al.*, 2020) ou pour prédire la structure d'un vaste nombre de protéines (Extance, 2021). Ils peuvent également servir à simuler des systèmes complexes plus efficacement que les outils classiques, avec des répercussions sur les prévisions météorologiques (Wolchover,



« Toutefois, réaliser l'équilibre entre encourager le développement de cette technologie — y compris les avantages sociaux et économiques connexes — et protéger les droits humains et s'assurer de l'utilisation éthique et responsable de l'IA tout au long de son cycle de vie se heurte à une divergence fondamentale. »

2018) et sur les simulations computationnelles dans de nombreux domaines (Ananthaswamy, 2021a). Dans un même temps, le développement des outils d'IA à des fins linguistiques peut modifier la façon dont les scientifiques interagissent avec la documentation scientifique en raison des outils qui synthétisent et résument les articles de recherche (Woolston et Perkel, 2020). Dans le domaine commercial, les bases de données de brevets peuvent être interrogées au moyen de l'IA à la recherche « d'art antérieur » afin de déterminer la nouveauté et donc la possibilité de breveter une invention (Helmers *et al.*, 2019). Les récentes extensions des outils linguistiques pourraient finir par permettre aux systèmes d'IA d'écrire du code informatique à partir d'invites textuelles de l'utilisateur (Chen *et al.*, 2021), offrant la possibilité d'automatiser les tâches de programmation scientifique (Hocky et White, 2022). Dans ce contexte, l'IA constitue

principalement un outil supplémentaire à la disposition des scientifiques et des ingénieurs dans le but de renforcer ou de compléter les pratiques classiques. Les exemples ci-dessus donnent néanmoins un indice des possibilités que l'IA joue un rôle plus à assertif dans la boucle de conception et de découverte.

Les mégadonnées ouvrent de nouvelles voies pour aborder la conception et la découverte à l'aide de l'IA, ce qui pourrait influencer sur le milieu et la culture de la recherche

Un des éléments de la conception et de la découverte est la recherche parmi les hypothèses ou les champs d'enquête pour dénicher de nouvelles connaissances ou, par la permutation des propriétés décrivant un système (ou un processus), en arriver à un concept adapté à l'accomplissement d'une fonction intéressante. Les problèmes de conception comportent généralement des variables d'entrée et des contraintes (p. ex. les matériaux dont un composant peut être fait ou les conditions dans lesquelles il fonctionne), et les combinaisons de variables et de contraintes définissent un énorme espace de recherche. L'IA continue à être exploitée par les scientifiques humains et les ingénieurs dans divers milieux pour

s'attaquer à ces problèmes « explosifs sur le plan combinatoire » (The Turing Institute, 2021), dans lesquels le nombre de variables ou la quantité de données — ou les deux — sont ingérables. Avec la croissance de la disponibilité des données sur les intrants et extrants des expériences et des problèmes de conception, les systèmes d'IA deviendront de plus en plus aptes à conseiller les scientifiques humains sur la façon de résoudre efficacement un problème ou d'atteindre un objectif en suggérant des processus que ces derniers pourraient ne pas envisager ou qui pourraient leur sembler contre-intuitifs.

Par exemple, les systèmes d'IA pourraient proposer des montages expérimentaux novateurs pour l'observation de certains phénomènes en physique quantique (Ananthaswamy, 2021b) ou optimiser les flux de réaction en chimie synthétique qui défient les meilleures pratiques accumulées au cours de décennies d'études empiriques (Jia *et al.*, 2019). Ces exemples illustrent comment les approches guidées par les données peuvent aujourd'hui être intégrées à presque toutes les étapes du cycle de vie scientifique (Ezer et Whitaker, 2019) et en quoi les systèmes d'IA peuvent passer d'un rôle d'outils utilisés par les humains pour analyser ou interpréter leurs données à celui de guides vers des hypothèses à explorer, des expériences à effectuer ou des concepts à mettre en application et à valider novateurs (encadré 2.1).

Parce que les mégadonnées sont un facteur habilitant essentiel, les efforts pour produire rapidement davantage de données scientifiques lisibles par l'ordinateur se multiplient. Les processus à haut débit dans les laboratoires ou les milieux industriels exploitent également l'automatisation pour générer d'immenses quantités de données qui alimenteront la boucle de conception et de découverte. Des laboratoires automatisés dédiés de ce genre voient le jour pour appuyer l'expérimentation à haut débit (NIST, 2020) et des installations sont modifiées dans le même but, en raison des progrès dans les technologies connexes (p. ex. robotique et optoélectronique) (Hatfield *et al.*, 2021). En génie et en fabrication, on pourrait créer et exploiter les mégadonnées pour les systèmes d'IA au moyen des techniques de jumeaux numériques, dans lesquelles le recueil de données à partir de capteurs incorporés dans un objet ou dans une ligne de production permet la conception d'une réplique entièrement numérique de cet objet ou du processus (Boschert *et al.*, 2018). Le système jumeau numérique évolue dynamiquement en parallèle avec son jumeau physique et les systèmes d'IA peuvent analyser les données qui en résultent à des fins de surveillance ou de prédiction et de simulation. Ce principe a des conséquences sur la conception et le fonctionnement de l'infrastructure physique et d'autres objets ou processus industriels complexes (Stevens *et al.*, 2020).

Encadré 2.1 Adoption précoce de l'IA pour la conception en science des matériaux

Certains domaines et certaines disciplines, comme la science des matériaux, ont rapidement décelé les avantages potentiels des techniques guidées par les données et la nécessité de créer et de synthétiser de grandes quantités de données (NASEM, 2018). Les matériaux nouveaux jouent un rôle crucial en rendant possibles de nombreuses applications technologiques, des biens de consommation à l'infrastructure (Wang *et al.*, 2020), mais il est difficile de prédire en quoi la composition d'un matériau est liée à sa fonction ou comment produire efficacement des matériaux arbitraires. Alors que les scientifiques et les chercheurs humains peuvent ajuster une variable à la fois pour vérifier ces liens, les systèmes pourraient effectuer la recherche très différemment (Stach *et al.*, 2021). Au lieu de compter uniquement sur l'intuition humaine, l'expérience ou la méthode d'essais-erreurs pour définir quels matériaux réels ou hypothétiques pourraient exceller dans une fonction donnée, les systèmes d'IA peuvent utiliser des corrélations trouvées dans les bases de données des propriétés des matériaux pour permettre aux scientifiques et aux ingénieurs de travailler à l'envers, c'est-à-dire partir de la fonction souhaitée, et éventuellement de concevoir des matériaux plus rapidement et plus efficacement.

Bien que l'intégration de l'IA soit plus avancée dans certaines disciplines et certains domaines, la disponibilité croissante des données peut permettre à d'autres disciplines des sciences naturelles, des sciences sociales et des sciences humaines — dont le lien avec l'IA est parfois moins intuitif — d'adopter les technologies d'apprentissage machine en laboratoire (The Royal Society, 2019). Gil *et al.* (2014) avancent que les domaines universitaires sont touchés par l'IA parfois à un point tel que les futurs progrès sont presque impensables sans la contribution de l'apprentissage machine. Le haut degré de collaboration internationale et institutionnelle dans le domaine de l'IA (Wu *et al.*, 2020) porte à croire que le développement continu de cette technologie accélérera la création de nouveaux partenariats dans le milieu de la recherche (chapitre 4).

La U.S. National Science Foundation met actuellement sur pied plusieurs instituts pour appuyer la collaboration interdisciplinaire et intersectorielle, qui utilisent l'IA comme plateforme pour régler des problèmes de grande importance sociétale (Bates, 2021). En fait, l'IA est reconnue comme un outil puissant de résolution des problèmes interdisciplinaires et son développement profite autant de cette collaboration que les autres domaines scientifiques profitent de l'utilisation de

l'IA (Kusters *et al.*, 2020). Les progrès continus dans la technologie et les mégadonnées se prêtent à modifier les flux de travaux de la recherche à l'échelle du laboratoire ainsi qu'au niveau des réseaux de recherche internationaux. Ces progrès auront des conséquences sur le financement de la recherche, la formation et l'infrastructure (chapitre 4).

La conception et la découverte autonomes par des machines avec peu ou pas d'apport humain sont réalisables, mais soulèvent de nombreuses questions et de nombreux défis

Les outils d'IA destinés aux sciences et au génie sont répandus, mais ils sont encore insuffisamment autonomes (Gil *et al.*, 2020). Bien que les systèmes d'IA aient prouvé leur capacité à soutenir les humains le long de la chaîne de conception et de découverte, leurs réalisations à ce jour ont nécessité des interventions et un apport humains conséquents (The Turing Institute, 2021).

Actuellement, l'intervention humaine est également requise pour transférer les capacités des systèmes d'IA d'une tâche à une autre.



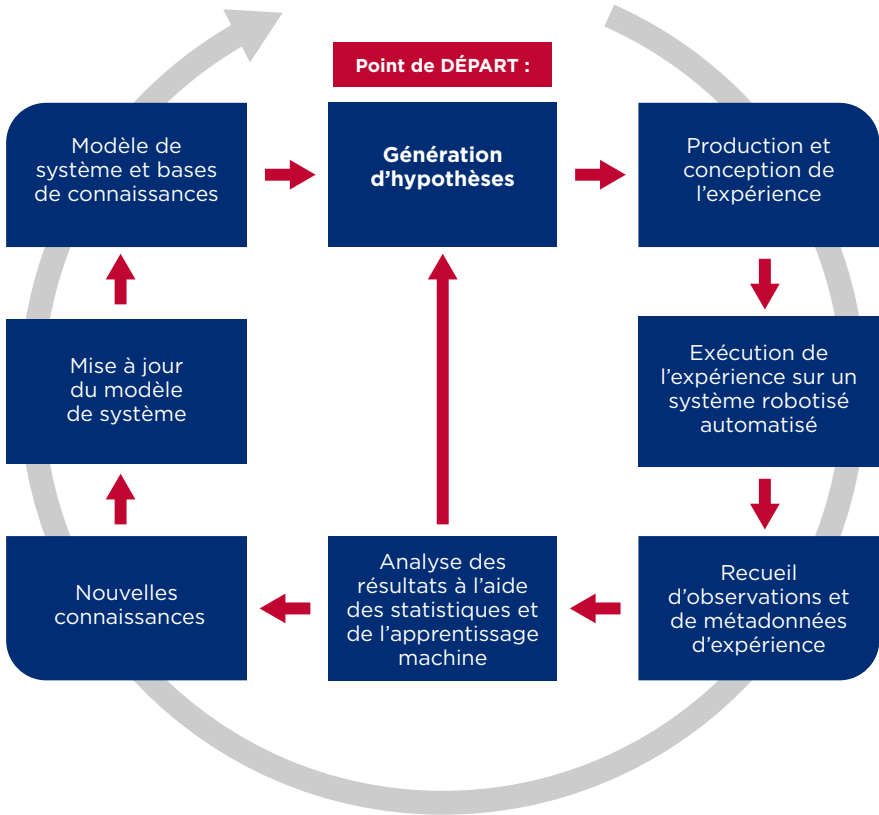
« l'IA est reconnue comme un outil puissant de résolution des problèmes interdisciplinaires et son développement profite autant de cette collaboration que les autres domaines scientifiques profitent de l'utilisation de l'IA. »

De nombreux systèmes sont extrêmement spécialisés et ne sont pas faciles à généraliser (The Turing Institute, 2021), il est donc difficile de les adapter ou de les transférer entre les utilisateurs, les disciplines et les secteurs. Ces systèmes offrent une capacité d'abstraction limitée et ont du mal à fournir un raisonnement scientifique soutenant les découvertes réalisées grâce à eux² (chapitre 3).

Des efforts sont en cours pour surmonter ces limitations et réduire l'ampleur de l'intervention humaine manifeste (figure 2.1) en augmentant le degré d'automatisation dans les laboratoires, notamment en chimie et en sciences biopharmaceutiques dans les secteurs universitaires

et privés (Sanderson, 2019; Mullin, 2021). Comme les limites humaines sont de plus en plus apparentes dans la gestion des forts volumes de données demandant un traitement rapide et exact, la demande de solutions d'IA croît, à tel point qu'un système d'IA entièrement automatisé est maintenant concevable (King *et al.*, 2009; Sparkes *et al.*, 2010; The Turing Institute, 2021). Comme l'IA prend une position de plus en plus centrale dans le cycle de la découverte, on s'attend à ce que le processus de demande de financement de la part des chercheurs change.

2 AlphaFold, un projet de DeepMind, a obtenu en 2021 un résultat qui fait date en prédisant la structure de protéines. Les prédictions servent d'intrants aux hypothèses formulées par les scientifiques humains, mais elles n'expliquent pas la biophysique sous-tendant le pliage des protéines ou d'autres liens pertinents sur le plan biologique entre la structure et la fonction (Extance, 2021).



Reproduit avec la permission de Springer Nature (Sparkes *et al.*, 2010)

Figure 2.1 Cycles de génération d'hypothèses et d'expérimentation automatisés

Exemple de flux de travail pour un système d'IA capable d'effectuer de manière autonome des recherches scientifiques pour aboutir à de nouvelles découvertes.

Cette figure fait partie d'un article publié sous licence par BioMed Central Ltd, un article en libre accès distribué selon les termes de l'attribution de licence Creative Commons (<https://creativecommons.org/licenses/by-nc-nd/2.0/ca/legalcode.fr>). Celle-ci autorise l'utilisation, la distribution et la reproduction sans restriction par tous moyens et sous tous formats, à condition que l'œuvre originale soit correctement citée.

Plusieurs pays déploient des programmes visant à réaliser des systèmes d'IA entièrement autonomes pour des applications en sciences et en génie dans le cadre du projet Grands Défis. Le gouvernement du Japon finance une initiative intersectorielle de production de systèmes robotisés commandés par l'IA, qui seraient capables de réaliser des principes et des solutions scientifiques percutants, entre autres objectifs bénéfiques pour la société (CSTI, 2020). Ailleurs, le Turing Nobel Challenge, lancé au Royaume-Uni, mais reposant sur la collaboration internationale et interdisciplinaire, cible la création d'un scientifique autonome capable de faire des découvertes susceptibles de valoir un prix Nobel d'ici à 2050 (Gil *et al.*, 2020; Kitano, 2021). Le système ainsi produit — soit un chercheur autonome — devrait être en mesure de démontrer



« Au lieu de remplacer purement et simplement les scientifiques humains, les systèmes d'IA pourraient être exploités pour offrir une nouvelle méthode de découverte scientifique, bien que de nombreuses conséquences demeurent incertaines. »

des connaissances d'expert, de communiquer les résultats et les découvertes aux humains et de produire des questions de recherche intéressantes (Gil *et al.*, 2020). Il formulerait des hypothèses, créerait les expériences nécessaires (Kitano, 2021) et pourrait déterminer les priorités d'exploration entre les divers moyens, y compris par l'estimation des impacts potentiels (chapitre 4).

Au lieu de remplacer purement et simplement les scientifiques humains, les systèmes d'IA pourraient être exploités pour offrir une nouvelle méthode de découverte scientifique, bien que de nombreuses conséquences demeurent incertaines. D'un point de vue social, les systèmes d'IA de cette nature entraîneraient une suppression de la main-d'œuvre humaine d'une ampleur inconnue et seraient

déployés dans des contextes sociaux extrêmement différents de ceux dans lesquels ils ont été développés (chapitre 5). Par exemple, des craintes liées à ce problème ont poussé à demander à ce que ces systèmes soient développés conjointement par des chercheurs et les communautés autochtones lorsqu'il est prévu que leur déploiement s'effectuera dans ces communautés (Lewis, 2020) (section 5.2).

D'autres questions potentielles sont plus ambiguës et floues, notamment comment accorder l'accès aux ressources d'IA de façon à éviter d'exacerber les inégalités existantes dans le système de recherche. Un chercheur autonome pourrait probablement rompre la dépendance de sentier de la recherche parce qu'il serait débarrassé de certaines mesures incitatives ou de certaines valeurs problématiques qui influencent l'orientation de la recherche faite par les humains (Kitano, 2021). Cependant, il serait toujours souhaitable qu'un tel chercheur

fonctionne en harmonie avec les valeurs humaines (The Turing Institute, 2021); après tout, la recherche consacrée au développement des scientifiques autonomes exigera invariablement d'importants investissements de la part des secteurs publics et privés. Pourtant, comme pour les véhicules autonomes ou les systèmes de recommandation reposant sur l'IA, l'intégration explicite des valeurs humaines dans les systèmes algorithmiques n'est pas toujours réalisable (Gibert, 2020), pas plus que ne sont simples les décisions quant aux valeurs à intégrer.

Enfin, bien que l'encodage dans les systèmes d'IA de l'expertise, de l'apprentissage et de l'expérience humains dans le domaine demeure un défi, (King *et al.*, 2018), l'inverse est également préoccupant. Pour que les humains utilisent efficacement les connaissances créées par l'IA, les découvertes doivent être interprétables, réalisées en temps opportun et pertinentes pour les scientifiques humains (The Turing Institute, 2021). Les défis épistémiques et éthiques qui sont analysés en profondeur dans les chapitres suivants représentent d'importants obstacles à surmonter pour en arriver à une boucle de conception et de découverte d'où les humains seraient absents.

2.2 Contexte canadien

Dès les années 1980, le CIFAR a mis sur pied des programmes de recherche en IA afin de soutenir la recherche fondamentale alors très peu populaire (Johnson, 2021). À l'époque, les chercheurs explorant l'apprentissage profond et l'apprentissage par renforcement étaient considérés comme à la périphérie de la communauté informatique. Cependant, ces démarches sont aujourd'hui largement perçues comme des réalisations importantes de l'IA d'origine canadienne (Smith, 2017). Au cours de la dernière décennie, la compétition en matière de développement de l'IA s'est accélérée partout dans le monde, à mesure que le domaine a dépassé le simple milieu de la recherche universitaire et s'est taillé une place dans la société. Dans ses efforts de développement de systèmes d'IA destinés aux sciences et au génie, le Canada aura à gérer ses forces et ses faiblesses intrinsèques en R-D dans un monde de l'IA de plus en plus multipolaire et dynamique.

Le milieu de l'IA actuel doit son établissement à des investissements publics à long terme et est aujourd'hui dominé par un petit nombre d'acteurs régionaux qui stimulent une croissance continue

La compétitivité du Canada en IA remonte à des investissements constants et précoces effectués dans le domaine, même quand cela n'était pas à la mode. Plusieurs personnages clés de la recherche en IA ont fait carrière au Canada en partie en raison du financement stable des chercheurs et de la recherche (Gherhes *et al.*, 2021). Plus récemment, le lancement en 2017 de la Stratégie pancanadienne

en matière d'intelligence artificielle a donné le coup d'envoi à un nouvel investissement majeur de fonds publics, qui a façonné le paysage actuel de l'IA au Canada. Cette stratégie, menée par CIFAR, a représenté la première stratégie nationale complète du genre (UNESCO, 2018) et offre un stimulus opportun permettant de tirer parti des jalons posés durant les décennies précédentes (Smith, 2017). Son objectif premier est de soutenir les activités dans trois principaux centres de recherche en IA : l'Alberta Machine Intelligence Institute (Amii), le Vector Institute et Mila (anciennement le Montreal Institute for Learning Algorithms) (Smith, 2017). Ces trois instituts sans but lucratif constituent d'importants pôles de R-D en intelligence artificielle au Canada, et accueillent des chaires de recherche et des savants et soutiennent des réseaux regroupant des établissements d'enseignement supérieur et l'industrie, dont de nombreuses firmes multinationales œuvrant dans le domaine (Chowdhury *et al.*, 2020). La Colombie-Britannique a aussi développé des forces en IA, en particulier en vision par ordinateur, et bénéficie d'un solide écosystème pour les entreprises en démarrage et d'un environnement favorable à la commercialisation des applications d'IA (Chowdhury *et al.*, 2020).

D'autres initiatives spécialisées axées sur l'application de l'IA aux problèmes des sciences ou du génie continuent de voir le jour. Le programme Défi « L'intelligence artificielle au service de la conception » du CNRC a pour but l'élaboration et l'exploitation de technologies et de compétences en IA qui accéléreront la découverte, la R-D et l'innovation. Il favorise la mise au point d'algorithmes, de méthodes et d'ensembles de données pour aider les ingénieurs, les chercheurs et les scientifiques à améliorer leurs procédés de conception et à multiplier leurs découvertes (CNRC, 2021). De son côté, l'Initiative des Supergrappes d'innovation du gouvernement fédéral, lancée en 2018, a pour but de créer des pôles de R-D partout au Canada dans de nombreux champs pour combler les lacunes en matière d'innovation et de technologie (Knubley, 2021). Située à Montréal, la supergrappe Scale IA cherche à appliquer l'IA aux produits commerciaux et aux chaînes d'approvisionnement, mais d'autres supergrappes sont également actives dans le domaine des applications d'IA ou dans des secteurs périphériques comme la robotique (Knubley, 2021). En 2021, l'Université de Toronto a lancé l'Acceleration Consortium, qui concerne la recherche sur les matériaux et réunit plus de 50 de meilleurs chercheurs de l'université et du monde (Kalvapalle, 2021). Le consortium se focalise en grande partie sur la détermination et la production de nouveaux matériaux prometteurs aux fins de développement technologique; cependant, il s'orientera également vers le renforcement des capacités et la recherche dans l'automatisation des laboratoires en général, la conception expérimentale assistée par l'IA et la formation de personnel hautement qualifié dans le domaine (Acceleration Consortium, 2021).

La stratégie de recherche sur l'IA du Canada a donné une impulsion et a établi une masse critique dans ses pôles, mais pour permettre la conception et la découverte en sciences et en génie, il faudra étendre la portée des activités actuelles

L'environnement de recherche en IA au Canada est spécialisé et présente des forces dans plusieurs secteurs, notamment l'apprentissage profond, la vision par ordinateur et l'apprentissage par renforcement (Chowdhury *et al.*, 2020)³. Les premières impressions donnent à penser que la Stratégie pancanadienne en matière d'intelligence artificielle a atteint plusieurs de ses buts, et le budget fédéral 2021 prévoyait de nouveaux fonds pour la prochaine décennie afin de renouveler cette stratégie (Chowdhury *et al.*, 2020; GC, 2021c). La contribution des chercheurs canadiens en IA et leur influence se reflète par le rang relativement élevé du pays dans les indicateurs traditionnels de publication et de dépôt de



« Selon le comité d'experts, l'écosystème d'IA canadien — jusqu'ici habitué à une croissance verticale — devra croître horizontalement au-delà de ses forces actuelles et des frontières physiques, disciplinaires et sectorielles pour tirer le meilleur parti des occasions de conception et de découverte en sciences et en génie. »

brevet (OCDE, 2021e). Le Canada se situe également en haut du Global IA Index (Tortoise Media, 2021), un classement basé sur une combinaison d'indicateurs pondérés, qui donne un aperçu agrégé de la capacité en IA d'un pays dans de multiples secteurs (Tortoise Media, 2020). La capacité de recherche du Canada dans certains sous-domaines, une des premières au monde, témoigne bien du soutien à long terme et ciblé apporté par le gouvernement à la recherche fondamentale en IA dirigée par les chercheurs (Gherhes *et al.*, 2021).

Néanmoins, certains éléments de la situation actuelle peuvent nuire à la compétitivité du Canada en ce qui concerne les applications d'IA pour les sciences et le génie. L'offre de programmes de formation et d'éducation dans les champs liés à l'IA y penche lourdement vers certains secteurs, c'est-à-dire la robotique et l'automatisation, alors

que dans d'autres pays elle porte sur un ensemble plus large de sous-domaines (OCDE, 2021b). Le Canada a été l'un des premiers à définir une stratégie nationale, mais cette stratégie était considérablement axée sur la recherche et les talents en IA, alors que des exemples internationaux plus récents accordent une plus grande importance aux acteurs au-delà de l'écosystème de recherche (Kung *et al.*, 2020). On s'inquiète également que l'approche suivie jusqu'ici, soit de concentrer le financement et les activités sur une poignée d'entités et de régions géographiques,

3 Les dépôts de demandes de brevet démontrent également une certaine spécialisation, en particulier dans des domaines tels que le traitement automatique des langues (ISDE, 2019a).

pourrait être problématique (Brandusescu, 2021). Toutefois, si d'après les données probantes, accroître la concentration du financement au-delà d'un certain seuil conduit à réduire le rendement marginal des investissements dans la recherche (Fortin et Currie, 2013; Lorsch, 2015; Mongeon *et al.*, 2016), un des principaux architectes du programme des supergrappes affirme que cette concentration répond aux préoccupations soulevées par certains intervenants selon lesquelles les ressources canadiennes en sciences, technologie et innovation sont trop éparpillées (Knubley, 2021).

Ces conflits sont inévitables, et bien que de nombreuses applications bénéficieront des forces des réseaux locaux, l'IA est une plateforme intrinsèquement portable. Les pôles devront donc être connectés aux activités dispersées géographiquement au Canada pour tirer efficacement parti de la capacité nationale de R-D. Selon le comité d'experts, l'écosystème d'IA canadien — jusqu'ici habitué à une croissance verticale — devra croître horizontalement au-delà de ses forces actuelles et des frontières physiques, disciplinaires et sectorielles pour tirer le meilleur parti des occasions de conception et de découverte en sciences et en génie.

2.3 Progrès en gouvernance et en politique

Malgré la rapide croissance du domaine, utilisation de l'IA comme plateforme technologique n'est pas actuellement assujettie à une réglementation exhaustive. Le potentiel de l'IA de causer des préjudices, d'exacerber les inégalités ou d'induire en erreur, entre autres risques, a motivé plusieurs tentatives de définir des lignes directrices concernant son développement éthique. De nombreux volets de la gouvernance, néanmoins, demeurent flous, que ce soient les problèmes et les différences de fond affichées dans les déclarations internationales ou les rôles et responsabilités concernant l'implantation et la mise en application des principes de gouvernance. Différentes approches commencent à prendre forme dans l'ensemble du spectre, qui s'étendent d'une réglementation « dure » par le biais de la législation à des régimes « souples », dans lesquels les praticiens s'engagent à respecter des normes et des codes de pratique volontaires. Toutefois, certaines applications d'IA liées aux sciences et au génie ne sont pas à proprement parler controversées; par exemple, la détection d'objets ou d'anomalie dans les données économiques n'a pas d'impact sur les droits humains. Néanmoins, et particulièrement à mesure que les systèmes d'IA deviennent plus autonomes, les progrès réalisés en gouvernance de l'IA décrits ci-dessous — ainsi que les autres tendances relatives à la gouvernance des données et à la culture de la recherche — en viendront inévitablement à définir l'espace dans lequel les chercheurs et les innovateurs créent, collaborent et fonctionnent.

Les multiples lignes directrices nationales et internationales sur l'éthique et la gouvernance de l'IA publiées à ce jour sont volontaires et leur adhésion est inégale

De nombreux ensembles de lignes directrices ont été proposés par des organismes gouvernementaux et non gouvernementaux dans le monde — fréquemment à la suite de consultations avec de multiples parties prenantes — pour promouvoir une recherche en IA responsable et centrée sur l'humain. La portée de ces lignes directrices est vaste, et la diversité de leurs recommandations souligne les oppositions qui existent entre la promotion de l'innovation et la nécessité de protéger les droits humains ou entre les considérations locales et internationales, entre autres (Gibert *et al.*, 2018). Ces oppositions découlent de la grande quantité d'intervenants dans l'IA aux priorités conflictuelles ou contradictoires. Par exemple, la Déclaration de Montréal IA responsable est une initiative canadienne de gouvernance de l'IA auquel de nombreux signataires ont adhéré au Canada comme à l'étranger, dont des organismes publics et privés ainsi que des individus (Déclaration de Montréal IA responsable, 2021). La Déclaration de Toronto, quant à elle, établit des lignes directrices pour empêcher le développement de systèmes d'apprentissage machine discriminatoire selon les lois internationales de protection des droits humains (Amnesty International et Access Now, 2018). À l'échelle nationale, en octobre 2021, 46 pays avaient officiellement souscrit aux principes établis par le Conseil sur l'intelligence artificielle de l'OCDE. Bien que ces lignes directrices procurent un ensemble de principes convenus à l'échelle internationale, elles ne sont pas juridiquement contraignantes et la vaste majorité des signataires (38 sur 46) sont eux-mêmes des pays membres de l'OCDE (OCDE, 2021c). L'adoption de lignes directrices ou leur adhésion est variable et correspond aux priorités des intervenants des gouvernements.

Les documents d'orientation actuels contiennent certains sujets communs. Ils mettent régulièrement l'accent sur le bien-être, l'autonomie, la justice, la confidentialité, le savoir, la démocratie et la responsabilité dans les cadres qu'ils proposent et qui recouvrent l'ensemble du spectre, de la réglementation souple à la réglementation dure (Gibert *et al.*, 2018). Un examen de 84 directives éthiques concernant l'IA a révélé que la priorité était accordée à un sous-ensemble de principes regroupant notamment la transparence, la justice et l'équité, la non-malfaisance, la responsabilité et la confidentialité. L'analyse a constaté une variabilité dans la priorisation et l'interprétation de ces principes, ainsi qu'un manque de clarté à propos des parties prenantes auxquelles ces directives devraient (ou ne devraient pas) s'appliquer (Jobin *et al.*, 2019). Bien qu'il n'y ait pas de consensus clair sur des lignes directrices universelles pour régir l'IA, les documents récents ont commencé à converger vers des éléments similaires (Fjeld *et al.*, 2020). Néanmoins, on continue à s'inquiéter que la voix de certains intervenants clés ne

soit toujours pas entendue. Quoique de nombreuses lignes directrices aient été élaborées par le biais de mécanismes de participation dans l'espoir de recueillir des idées d'une vaste diversité de parties prenantes qui seront touchées par l'IA, les individus ou les organisations des démocraties occidentales libérales ont tendance à être surreprésentés (Gibert *et al.*, 2018). Dans leur examen des directives éthiques, Jobin *et al.* (2019) ont constaté une sous-représentation au niveau mondial de l'Afrique, de l'Amérique du Sud et centrale et de l'Asie centrale. Une plus grande démocratisation et une participation plus inclusive dans l'IA permettraient l'ajout de nouvelles priorités aux principes existants (Fjeld *et al.*, 2020).

La manière de mettre en pratique ces directives éthiques n'est toutefois pas précisée dans la plupart des documents, pas plus que les cadres de résolution des conflits entre les principes et de respect de ces lignes directrices (Jobin *et al.*, 2019; Whittlestone *et al.*, 2019). À ce titre, il existe de nombreuses lacunes dans la mise en œuvre des déclarations sur l'IA éthique ou responsable (Whittlestone *et al.*, 2019; Fjeld *et al.*, 2020) et les approches souples de la réglementation sont encore la norme mondiale (OCDE, 2021b). Les individus et les organisations qui appliquent activement l'IA aux sciences et au génie doivent être conscients de la fluidité de la gouvernance de ce domaine aux échelles internationale, nationale et sous-nationale. Morley (2021) tente de combler le fossé entre les principes et les pratiques au moyen d'une typologie visant à aider les développeurs en éthique appliquée à chaque étape de la filière de développement de l'apprentissage machine. Bien que ces travaux portent uniquement sur l'apprentissage machine, la recherche peut être applicable à d'autres catégories d'IA. L'application des principes éthiques à l'utilisation de l'IA en sciences et en génie est examinée plus en détail et à la section 3.2.

Les applications d'IA accroîtront les liens et les interdépendances au sein de la communauté de l'IA et du milieu des sciences et du génie dans son ensemble, ce qui aura des conséquences sur les politiques et la gouvernance

Les systèmes d'IA ne se limitent pas aux algorithmes qui définissent leur fonctionnement. Ces systèmes sociotechniques reposent également sur les données, sur une technologie complémentaire (p. ex. les capteurs) et sur les contextes sociaux dans lesquels ils sont développés et déployés (OCDE, 2019b). Les efforts et les stratégies de promotion de l'IA doivent tenir compte des croisements entre les politiques élaborées dans des domaines proches.

Dans la communauté universitaire, la modification des structures d'incitation et de récompense concernant la recherche scientifique influe sur l'évolution de l'IA en sciences et en génie, par exemple par l'allocation des fonds de recherche. Les organismes de financement de la recherche continuent à faire progresser les méthodologies d'évaluation de la recherche et des chercheurs (GRC, 2021), ce qui aura de nombreuses conséquences si les humains sont retirés de la boucle (chapitre 4). Par exemple, l'initiative Recherche et innovation responsables de l'Union européenne (UE) cherche à changer la culture de la recherche en promouvant des approches de la science et de l'innovation qui soient éthiques, inclusives et ouvertes (RRITools, s.d.).

De plus, l'OCDE a recommandé de favoriser les données ouvertes pour promouvoir l'accès aux données dans tous les secteurs, dans le but de faciliter le développement de l'IA (OCDE, 2019a). Par conséquent, les investissements dans l'infrastructure des données et dans l'élaboration de politiques sur ces mêmes données sont également essentiels pour promouvoir l'adhésion à l'IA et son adoption. Ces investissements pourraient inclure l'Internet des objets, l'accès aux réseaux à large bande et le calcul et le stockage de haute performance (OCDE, 2021b). L'évolution des cadres politiques concernant les données influencera les activités d'IA en sciences et en génie, même si ces cadres ne ciblent pas principalement l'innovation dans ces domaines, comme c'est le cas de la National Data Strategy, au Royaume-Uni (Gouv. du Royaume-Uni, 2020). L'accès aux données est étudié plus en détail au chapitre 6.

L'espoir que l'IA permettra de s'attaquer à d'autres objectifs politiques mondiaux aura également une incidence sur la cadence de son développement et sur la priorisation des applications, particulièrement en sciences et en génie.

Répercussions de l'IA sur la conduite de la recherche

- 3.1 Garantir l'intégrité scientifique dans la recherche utilisant l'IA
- 3.2 Utilisation éthique de l'IA dans la recherche en sciences et en génie
- 3.3 Les répercussions de l'IA sur les valeurs et sur les pratiques sociales en sciences

Constatations du chapitre

- Le domaine de la recherche sur l'IA fait actuellement face à des problèmes de reproductibilité, qui pourraient compliquer son utilisation dans d'autres disciplines scientifiques.
- Certains types populaires de systèmes d'IA fonctionnent comme des boîtes noires, ce qui fait qu'il est parfois difficile, voire impossible, d'expliquer comment les résultats sont obtenus; ce problème peut nuire à l'explication et à la compréhension scientifiques et miner la crédibilité des découvertes scientifiques produites par la machine.
- La précision des résultats générés par l'IA est déterminée, en partie, par la qualité des données avec lesquelles le système a été entraîné. Les systèmes d'IA peuvent produire des résultats inexacts ou faussés à cause de biais dans les ensembles de données d'apprentissage et de problèmes de généralisation des données d'apprentissage aux nouvelles données.
- Les cadres éthiques actuels de conduite de la recherche ne portent pas sur la complexité que l'IA introduit dans les principes classiques de l'éthique de la recherche, comme les participants humains aux recherches et le consentement éclairé.
- L'intensification de l'utilisation de l'IA pourrait modifier l'importance relative des différentes valeurs de la recherche en sciences et en génie ainsi que leur dynamique sociale sous-jacente.

L'intensification de l'utilisation de l'IA dans la recherche en sciences et en génie influera probablement sur la nature et la pratique de la recherche scientifique, ainsi que sur la compréhension qu'ont les chercheurs de l'intégrité scientifique, de la conduite éthique de la recherche et des valeurs et pratiques sociales de la science. L'IA interprétable et explicable sera essentielle en sciences et en génie pour permettre aux scientifiques humains d'apprendre des découvertes et des conceptions de l'IA, de réduire les risques et d'encourager la confiance et la transparence dans la recherche. Il pourrait également être nécessaire d'élaborer de nouveaux cadres d'éthique et de responsabilité pour la recherche effectuée au moyen de l'IA, afin de répondre aux nouveaux défis liés à des questions telles que le consentement éclairé, les préjudices sociaux découlant de l'IA et l'utilisation et la réutilisation des données. En outre, en plus des valeurs épistémiques et éthiques, l'IA peut aussi transformer les valeurs qui régissent la pratique sociale de la science, comme celles qui sous-tendent la validation, la provenance et la diffusion des résultats scientifiques.

3.1 Garantie de l'intégrité scientifique dans la recherche utilisant l'IA

L'augmentation de l'utilisation de l'IA dans la recherche en sciences et en génie apporte de nouveaux défis épistémiques et éthiques en ce qui concerne l'intégrité de la recherche. L'intégrité scientifique comprend une grande variété de facettes, comme la gestion des conflits d'intérêts, la reconnaissance des contributeurs et de la paternité et l'utilisation des fonds de recherche (CAC, 2010). Cette section porte sur les caractéristiques de l'intégrité scientifique qui créent des défis uniques dans le contexte de l'IA appliquée aux sciences et au génie : la reproductibilité, l'interprétabilité, l'exactitude et le biais et la gestion des données. Plus précisément, les découvertes produites par l'IA peuvent être difficiles à valider en raison d'obstacles à la reproductibilité indépendante ainsi que du manque d'explicabilité ou d'interprétabilité de certains types de systèmes d'IA et de leurs résultats. Par conséquent, il est parfois difficile de vérifier l'exactitude des découvertes produites par l'IA et de garantir qu'elles sont exemptes de biais injustifiés. S'attaquer à ces problèmes exigera la transparence et la reddition de comptes dans les pratiques de recherche et dans la gestion et la gestion des données.

3.1.1 Reproductibilité

La reproductibilité est un des moyens pour la communauté scientifique de valider l'exactitude des découvertes ou des nouvelles constatations et est considérée comme une des « caractéristiques d'une bonne démarche scientifique » [traduction libre] (NASEM, 2019). Au sens large, la reproductibilité est la capacité des chercheurs indépendants à obtenir des résultats identiques (ou similaires) à ceux d'une étude précédente à partir de méthodes identiques (ou similaires), démontrant ainsi la validité de cette étude⁴. Bien que l'IA ait la possibilité d'améliorer la reproductibilité en sciences (King *et al.*, 2009), son intégration dans le processus scientifique peut également créer des problèmes (Carter *et al.*, 2019; The Royal Society, 2019). Certaines personnes prétendent d'ailleurs que la recherche sur l'IA se heurte à d'importants écueils en matière de reproductibilité (Hutson, 2018; Heaven, 2020), qui pourraient nuire à d'autres domaines de recherche en sciences et en génie qui ont recours à l'IA.

4. Les termes tels que *reproductibilité* et *réplicabilité* sont parfois définis et utilisés de façon distincte et même contradictoire dans différentes disciplines (Fiddler et Wilcox, 2018; NASEM, 2019). En fait, même à l'intérieur du champ de la recherche sur l'IA, ils peuvent avoir des significations opposées selon la source; comparer par exemple Gundersen et Kjensmo (2018) et Nagarajan *et al.* (2019), d'une part, et NASEM (2020) et Carter *et al.* (2019), d'autre part. Toutefois, dans ce rapport, le terme *reproductibilité* est simplement utilisé comme un terme général recouvrant une grande variété de situations, de manière interchangeable avec *réplicabilité*.

L'IA a la capacité d'améliorer la reproductibilité en sciences

L'IA a la capacité d'améliorer la reproductibilité en sciences en permettant la description et l'enregistrement d'expériences de façon plus détaillée et avec une plus grande clarté sémantique que les humains (King *et al.*, 2009, 2018). Comme les renseignements sur l'installation et les procédures expérimentales — c.-à-d. les « métadonnées expérimentales » (King *et al.*, 2009) — sont automatiquement enregistrés par les systèmes d'IA avec une fidélité parfaite grâce au langage logique formel, les chercheurs autonomes mus par l'IA (ou « scientifiques robots ») peuvent parfaitement recréer tous les volets pertinents d'une expérience. En revanche, non seulement les humains peuvent faire des erreurs ou des omissions quand ils enregistrent ces renseignements, mais ces activités prennent aussi du temps et exigent de connaître les normes de production de rapport. De plus, contrairement à l'IA, les humains enregistrent généralement les métadonnées expérimentales en langages naturels, qui peuvent introduire des ambiguïtés et des imprécisions, ce qui empêche la reproductibilité (King *et al.*, 2018).



« Par conséquent, il est parfois difficile de vérifier l'exactitude des découvertes produites par l'IA et de garantir qu'elles sont exemptes de biais injustifiés. S'attaquer à ces problèmes exigera la transparence et la reddition de comptes dans les pratiques de recherche et dans la gestion des données. »

L'amélioration de la transparence est nécessaire pour faciliter la reproductibilité de la recherche en IA

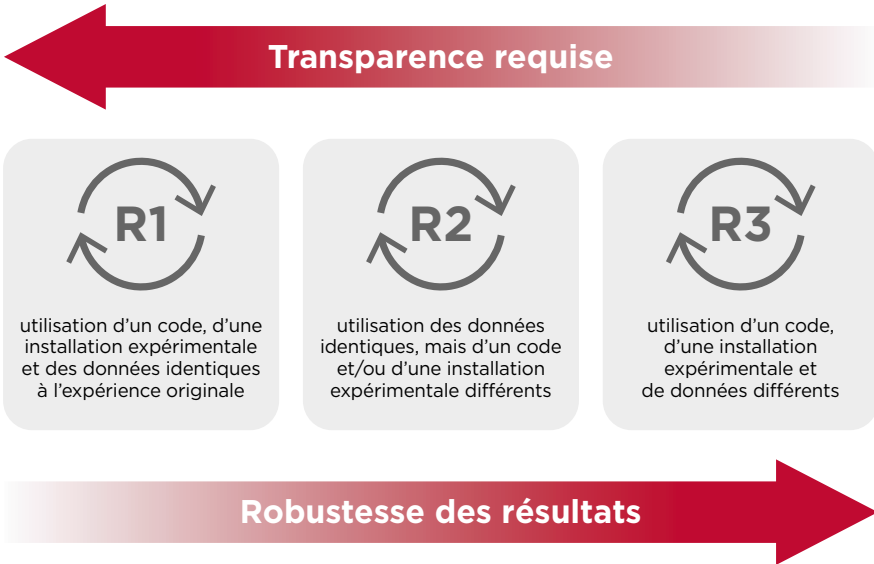
Un des principaux défis pour la reproductibilité scientifique en IA est le manque de transparence des chercheurs, qui ne fournissent pas toujours des renseignements suffisants ou assez détaillés pour permettre à leurs confrères et consœurs de répliquer leurs constatations. Pour que les résultats générés par la machine soient reproductibles, les chercheurs doivent fournir des renseignements sur le code, les données et l'infrastructure de calcul avec lesquels l'expérience a été réalisée (The Royal Society, 2019; Haibe-Kains *et al.*, 2020; Heaven, 2020). Malgré cela, un examen des publications effectuées par Gundersen *et al.* (2018) a révélé que seulement environ 8 % des articles présentés lors des grandes conférences sur l'IA entre 2013 et 2016 contenaient le code et que seulement environ un tiers contenaient les ensembles de données d'essai.

Simplement fournir le code est toutefois souvent insuffisant pour garantir la reproductibilité de la recherche sur l'IA parce qu'un même code peut produire différents résultats quand il est exécuté dans conditions dissemblables, par exemple en ce qui concerne le matériel ou les compilateurs, l'environnement logiciel, les

hyperparamètres (c.-à-d. les variables qui contrôlent le processus d'apprentissage, telles que l'architecture du réseau neuronal et le facteur d'apprentissage) ou les valeurs d'amorçage aléatoires (c.-à-d. les facteurs utilisés pour initialiser la pondération des connexions d'un réseau neuronal) (Henderson *et al.*, 2018; Hutson, 2018; Nagarajan *et al.*, 2019). Par conséquent, la reproductibilité exige généralement des chercheurs qu'ils fournissent des renseignements supplémentaires sur les conditions dans lesquelles leurs modèles sont entraînés et mis au point. Fournir l'accès aux données à partir desquelles les modèles d'IA ont été élaborés est également essentiel à la reproductibilité (Haibe-Kains *et al.*, 2020) parce qu'un même code peut fournir des résultats différents quand l'apprentissage s'effectue sur des ensembles de données différents (Hutson, 2018). Les données qu'il est utile de partager sont les données d'apprentissage, de validation et d'essai et les résultats (Gundersen *et al.*, 2018). Les principaux obstacles à la transparence du partage des données sont la protection des renseignements exclusifs, les problèmes liés à l'attribution de licence et les craintes concernant la confidentialité (lorsque les données concernent des sujets humains) (Haibe-Kains *et al.*, 2020; Heaven, 2020).

La reproductibilité dans la recherche sur l'IA peut s'obtenir selon divers degrés

Il est possible de distinguer différents degrés de reproductibilité dans la recherche sur l'IA. Au sens large, les chercheurs peuvent tenter de reproduire les résultats identiques (ou similaires) à ceux de l'expérience originale au moyen : (R1) d'un code, d'une installation expérimentale et de données identiques, (R2) de données identiques, mais avec un code ou une installation expérimentale différent ou (R3) d'un code, d'une installation expérimentale et de données différents (Gundersen *et al.*, 2018) (figure 3.1). Fait notable, c'est la situation R1 qui nécessite le plus de transparence de la part des chercheurs en IA parce qu'elle nécessite une documentation précise (ou l'accès à ces éléments) du code original, de l'installation expérimentale et des ensembles de données. En revanche, c'est la situation R3 qui requiert le moins de transparence parce qu'elle n'a besoin que d'une description suffisamment détaillée des objectifs et de la méthode. Cependant, les résultats de R1 sont également les moins intéressants ou généralisables, car ils ne font que confirmer que les résultats de l'expérience initiale peuvent être reproduits dans des conditions extrêmement précises et strictes. R3 est la situation la plus intéressante, car elle prouve que les mêmes résultats peuvent être produits à l'aide d'une démarche et de données différentes (Gundersen *et al.*, 2018).



Adapté avec permission : Tous droits réservés © 2018, Association for the Advancement of Artificial Intelligence (Gundersen *et al.* 2018)

Figure 3.1 Degrés de reproductibilité de la recherche en IA

La reproductibilité peut prendre différentes formes et assurer différentes fonctions selon la discipline

Il est important de différencier la reproductibilité dans la *recherche* en IA de la reproductibilité dans d'autres champs de la recherche scientifique qui *utilisent* l'IA. La reproductibilité peut prendre différentes formes dans différentes disciplines scientifiques et peut avoir divers objectifs scientifiques et assurer diverses fonctions épistémiques (Leonelli, 2018). Par exemple, les tentatives pour reproduire la recherche concernant les organismes modèles en psychologie et en neurosciences expérimentales peuvent ne pas avoir pour but de reproduire directement des résultats antérieurs, mais plutôt de déterminer les sources des variations entre les installations expérimentales susceptibles d'influer sur l'interprétation des données produites par l'expérience. La recherche concernant des objets ou des matériaux rares, périssables ou inaccessibles — comme des spécimens organiques ou des restes archéologiques ou paléontologiques uniques — n'est souvent pas directement reproductible parce que l'unicité et l'impossibilité de duplication des objets ou des matériaux sont cruciales pour leur valeur de témoignage. De même, la recherche en médecine, en histoire et en sciences sociales repose souvent sur l'observation plutôt que sur l'expérimentation, la reproductibilité dépend donc de l'expertise d'observateurs qualifiés (Leonelli, 2018). En résumé, comme les objectifs et le rôle

de la reproductibilité varient selon la discipline scientifique, ils peuvent également varier en fonction de la discipline dans laquelle l'IA est utilisée. Il n'est donc pas toujours approprié ou utile d'appliquer un seul ensemble de normes de reproductibilité à la recherche en IA et dans d'autres domaines de recherche qui utilisent l'IA.

Plusieurs initiatives promeuvent une meilleure reproductibilité dans la recherche utilisant l'IA

Il existe plusieurs plateformes en ligne de recherche en IA plus transparentes et reproductibles, comme GitHub, qui permet de partager du code, ou TensorFlow, qui sert à partager des modèles et des cadres d'IA (Haibe-Kains *et al.*, 2020). Isdahl et Gundersen (2019) ont réalisé une analyse comparative afin de déterminer les plateformes d'apprentissage machine en ligne — comme OpenML, BEAT ou Floydhub — qui favorisent le plus la reproductibilité et par quels moyens. Dans les cas où l'ensemble de données n'est pas partageable, il existe d'autres solutions, comme fournir des renseignements sur les techniques de recueil de données et sur les étiquettes des données afin que d'autres chercheurs puissent construire des ensembles de données semblables, ou encore permettre à des vérificateurs indépendants d'accéder aux données afin de vérifier les résultats (Haibe-Kains *et al.*, 2020). Gebru *et al.* (2021) proposent des *feuilles de calcul pour les ensembles de données*, qui pourraient faciliter la reproductibilité de la recherche sur l'apprentissage machine en fournissant des métadonnées détaillées sur les ensembles de données. Elles incluent le motif de la création de ces ensembles de données, leur composition, leur processus de recueil, leur processus de prétraitement et d'étiquetage, leurs usages recommandés (y compris en matière de distribution et de mise à jour) et toutes les autres fonctions pertinentes (Gebru *et al.*, 2021). Pour améliorer la répliquabilité de la recherche en IA, des ingénieurs d'IBM Research ont mis au point un outil d'IA qui recrée le code source non publié à partir de la description qui en est faite dans un article (Sethi *et al.*, 2018).

Carter *et al.* (2019) proposent plusieurs solutions possibles aux défis du partage du code et des données. Elles incluent (i) un « examen virtuel » vidéo partageable, à peu près équivalent à une inspection réglementaire, dans lequel le code est examiné en détail et démontré à l'écran; (ii) la création d'un « environnement de calcul protégé », dans lequel les examinateurs ont accès aux données et à la modélisation en lecture seulement et dans une forme non exportable et (iii) la mise à la disposition des examinateurs des données, du code et de l'infrastructure de calcul par le biais d'un accord de licence approprié, qui protège juridiquement la sécurité des données, la confidentialité et la propriété intellectuelle.

Les éditeurs universitaires, les réviseurs et les organisateurs de conférences ont également un rôle à jouer dans la promotion de la reproductibilité. Par exemple, la conférence sur l'IA NeurIPS a récemment établi une liste de contrôle des exigences de communication que les chercheurs doivent remplir quand ils soumettent un article afin d'accroître la transparence et la reproductibilité (NeurIPS, 2021). De même, l'atelier JOURNE, qui s'est tenu au printemps 2021 dans le cadre de la Conference on Machine Learning and Systems (MLSys), portait sur l'amélioration de la transparence dans la recherche sur l'apprentissage machine et offrait des réflexions essentielles sur les résultats négatifs, les conclusions intermédiaires et le développement d'idées de recherche afin de corriger les biais qui incitent à ne publier que les bons résultats (JOURNE, 2021). Des initiatives telles que celles-ci sont importantes pour faciliter la reproductibilité parce que la recherche actuelle et les incitations à la publication (en IA, mais également dans d'autres champs) ne sont généralement pas propices à la reproductibilité à cause des pressions à publier rapidement, du manque de temps pour tester les algorithmes dans différentes conditions, du manque d'espace pour documenter toutes les configurations de l'installation expérimentale essayées et du fait que la publication des échecs de réplication est dissuadée (Hutson, 2018; NASEM, 2019).

3.1.2 L'interprétabilité, l'explicabilité et le problème de la boîte noire

Un des attributs les plus prometteurs de l'IA est sa capacité à déceler dans les données les modèles invisibles aux humains, ce qui permet des découvertes scientifiques (Samek et Müller, 2019; The Royal Society, 2019). Cependant, plusieurs méthodes d'IA populaires — comme celles basées sur les réseaux neuronaux d'apprentissage profond — sont souvent en mesure de produire des résultats extrêmement précis, mais fonctionnent encore parfois comme des boîtes noires et leurs utilisateurs, et même leurs concepteurs, ne peuvent pas expliquer comment les résultats ont été générés ou sur quelles fonctions des données ils sont fondés (Knight, 2017; The Royal Society, 2019). Bien que dans certains cas, l'exactitude seule puisse être suffisante pour les progrès scientifiques, le but de la science est ultimement *l'explication* et la *compréhension* (The Royal Society, 2019). Comme le formulent Samek et Müller (2019) :

En sciences [...] expliquer et interpréter les caractéristiques utilisées par le système d'IA pour prédire est souvent plus important que la prédiction elle-même parce que cela révèle des renseignements sur les mécanismes biologiques, chimiques ou neuronaux et peut ouvrir la porte à de nouvelles perspectives scientifiques [Traduction libre].

Pour ces raisons, les chercheurs dans certains champs scientifiques en sont venus à préférer les modèles d'IA qui sont interprétables et explicables, même au détriment de l'exactitude prédictive (Lapuschkin *et al.*, 2019). En fait, il peut y avoir un conflit entre l'explicabilité et l'exactitude prédictive, certains des systèmes d'IA les plus exacts sur le plan prédictif étant parmi les plus opaques et les moins explicables, alors que les plus transparents et explicables offrent parfois une faible exactitude (Gunning et Aha, 2017). Cependant, l'idée qu'il y a nécessairement un compromis à faire entre l'interprétabilité et l'exactitude a été remise en question par Rudin et Radin (2019), qui ont étudié des exemples de systèmes d'IA interprétables dont l'exactitude est équivalente (ou meilleure) à celle des systèmes boîtes noires dans des domaines tels que la justice pénale, les soins de santé et la vision par ordinateur.



« Il a été suggéré que la complexité et l'opacité inhérentes à certains types de systèmes d'IA soulevaient la question de savoir si les résultats scientifiques produits par ces systèmes pouvaient être (ou pourraient devenir) inintelligibles pour l'humain »

Il y a plusieurs raisons de préférer des modèles interprétables aux boîtes noires dans le contexte de l'IA pour les sciences et le génie. Avec les systèmes d'IA boîte noire, il peut être plus difficile de déterminer les cas dans lesquels les prédictions sont correctes, mais sont obtenues à partir de corrélations factices (Samek et Müller, 2019). Par exemple, un tel système a réussi à classer avec exactitude les images d'un ensemble de données de référence largement utilisée portant sur des objets tels que des bateaux, des avions et des chevaux, mais une analyse subséquente a permis de découvrir qu'il était parvenu à cette exactitude prédictive, en partie, en raison de

la fréquence de filigranes de la source sur les images de chevaux et d'un genre particulier de motif de bordure sur les images d'avions (Lapuschkin *et al.*, 2019). Quand le système d'IA est une boîte noire, il peut être presque impossible de déceler de telles corrélations factices. La science vise généralement à comprendre les relations causales dans le monde naturel, un genre de relation que les associations apprises par l'IA ne reflètent pas nécessairement (Lipton, 2018). De plus, l'aptitude à interpréter les résultats d'un modèle d'IA permet un processus itératif de raffinement et de traitement qui améliore l'exactitude du modèle, ce qui pourrait aussi mener à une meilleure exactitude globale qu'avec des modèles boîte noire (Rudin et Radin, 2019). L'interprétabilité des modèles d'IA est importante pour déceler les résultats potentiellement injustes ou discriminatoires (section 5.2) et pour garantir la confidentialité des données (Lipton, 2018; Rudin, 2019).

Il a été suggéré que la complexité et l'opacité inhérentes à certains types de systèmes d'IA soulevaient la question de savoir si les résultats scientifiques produits par ces systèmes pouvaient être (ou pourraient devenir) inintelligibles pour l'humain (Nickles, 2018). C'est-à-dire que les conclusions produites par l'IA pourraient être ce que Humphreys (2009) qualifie d' « opaque sur le plan épistémique », à un point tel que les humains n'auraient pas la capacité cognitive de bien comprendre le processus par lequel le système d'IA en est arrivé à ces conclusions (Leonelli, 2020). Ce principe est lié au concept d'*apprentissage machine ultrafort*, qui fait référence à la capacité de générer une hypothèse (i) que le système peut enseigner à un humain et (ii) par conséquent, qui permet à un humain d'améliorer son rendement prédictif sur une tâche par rapport à celui d'un autre humain qui a étudié les données d'apprentissage seules (Muggleton *et al.*, 2018). L'apprentissage machine ultrafort peut être compris comme une mesure de la compréhension humaine et de sa capacité cognitive concernant l'IA et de la transférabilité de nouvelles capacités de raisonnement d'IA à l'humain.

L'IA explicable (XAI) vise à créer des systèmes d'IA transparents mieux compréhensibles par les humains

Ces dernières années, le concept de XAI a énormément gagné en popularité et plusieurs ateliers et programmes ont été organisés sur le sujet (Adadi et Berrada, 2018). Par exemple, en 2017, la U.S. Defense Advanced Research Projects Agency (DARPA) a lancé son programme de XAI dans le but de créer des techniques d'apprentissage machine, ou de modifier des techniques existantes, afin de permettre l'explicabilité et la compréhension (Gunning et Aha, 2017). Bien qu'il n'existe pas de norme ou de définition généralement acceptée de ce que constitue la XAI (puisque le terme fait davantage référence à un mouvement qu'à un concept précis), il est largement reconnu qu'il est nécessaire à la confiance, à la compréhension et à la gestion efficace des résultats d'IA parce qu'il permet de comprendre les forces et les faiblesses d'un système d'IA et de détecter les erreurs ou biais potentiels (Gunning et Aha, 2017; Adadi et Berrada, 2018).

Il existe une variété de méthodes et de techniques utilisées pour rendre les modèles d'IA plus interprétables (voir Adadi et Berrada (2018), Samek et Müller (2019) et Lipton (2018)). Ces techniques se différencient par le fait qu'elles sont intrinsèquement interprétables ou seulement interprétables par explication ultérieure, que les interprétations qu'elles fournissent sont globales (elles expliquent la logique de l'ensemble du modèle) ou locales (elles expliquent seulement les résultats particuliers produits par le modèle) et qu'elles sont propres à certains types de modèle d'IA ou s'appliquent à toutes les classes de modèles (Adadi et Berrada, 2018).

3.1.3 Exactitude et biais

Il est essentiel pour l'utilisation de l'IA en sciences et en génie de garantir que ses résultats sont exacts et non biaisés. Les techniques d'IA peuvent être sujettes au *surapprentissage*, problème selon lequel un algorithme d'apprentissage machine peut produire des résultats exacts quand il est appliqué à ses données d'apprentissage, mais des résultats inexacts quand il est utilisé avec de nouvelles données (Ying, 2019). Ce problème peut être dû à plusieurs facteurs, tels qu'un ensemble de données trop restreint, un faible ratio signal-bruit dans l'ensemble de données (c.à.d. quand l'ensemble de données comporte un nombre important de corrélations factices qui rend difficile de détecter les corrélations significatives (Lehr et Ohm, 2017)) ou simplement un algorithme mal étalonné. Il existe une variété de solutions, qui s'étendent de l'amélioration ou de l'agrandissement des données d'apprentissage à la sélection de différents modèles, en passant par l'optimisation des hyperparamètres du modèle. D'autres solutions sont également possibles, comme l'*apprentissage ensembliste*, dans lequel des prédictions d'algorithmes entraînés différemment sont combinées pour produire des résultats plus exacts, et la *validation croisée*, dans laquelle le même algorithme est appliqué à différents sous-ensembles de données pour permettre la comparaison. Dans le cas des systèmes d'IA basés sur des réseaux neuronaux, les développeurs peuvent utiliser des approches propres à la technique telle que l'*abandon*, méthode selon laquelle des nœuds du réseau sont abandonnés aléatoirement pendant l'apprentissage (voir Bejani et Ghatee (2021) et (Ying, 2019) pour un examen de différentes solutions au surapprentissage).

L'exactitude des résultats produits par un algorithme d'IA est déterminée, en partie, par la qualité des données avec lesquelles le système a été entraîné. Les gros ensembles de données qui n'ont pas été soumis à un examen et à des contrôles de la qualité minutieux présentent plus de risques de contenir des exactitudes et des biais non pris en considération qui pourraient se traduire par des résultats trompeurs ou incorrects (Leonelli, 2020). De plus, les biais humains dans la construction des ensembles de données aux fins de recherche scientifique peuvent nuire au modèle d'IA utilisé pour faire avancer cette recherche. Par exemple, Jia *et al.* (2019) ont prouvé qu'un modèle d'IA entraîné sur petit ensemble de données concernant des réactions chimiques sélectionné aléatoirement surpassait un modèle entraîné sur un ensemble plus gros, mais sélectionné par l'humain, pour la prédiction des résultats de la réaction. La raison est que l'ensemble de données choisi par l'humain était influencé notamment par la popularité relative de certains réactifs ou de certaines conditions de réaction et par la fréquence à laquelle ils apparaissent dans les publications scientifiques (Jia *et al.*, 2019).

Cet exemple prouve à quel point les biais dans les modèles d'IA peuvent découler de décisions concernant le recueil et la curation des données d'apprentissage. Ce processus n'est jamais entièrement neutre ou objectif parce que les chercheurs doivent décider quelles données seront recueillies (et comment) et lesquelles seront rejetées ou ignorées (boyd et Crawford, 2012). De plus, les biais dans les données



« Il est crucial d'envisager les formes de biais associés aux systèmes d'IA conjointement avec les systèmes de reddition de comptes liés aux décisions concernant la gouvernance, la gestion et le traitement des données. »

peuvent résulter de facteurs pratiques qui ne sont pas purement épistémiques, tels que ceux liés au conditionnement, au stockage et au partage des données (Leonelli, 2020). Les chercheurs doivent également prendre diverses décisions subjectives, par exemple sur le choix des variables cibles, si et comment étiqueter les données d'apprentissage, comment traiter les valeurs aberrantes, comment partitionner les données pour l'essai, quels algorithmes choisir et comment mettre au point le modèle (Lehr et Ohm, 2017; Selbst et Barocas, 2018), lesquelles peuvent toutes introduire un autre biais.

Il est crucial d'envisager les formes de biais associés aux systèmes d'IA conjointement avec

les systèmes de reddition de comptes liés aux décisions concernant la gouvernance, la gestion et le traitement des données. Du point de vue du comité d'experts, il est impératif de continuer à superviser qui prend de telles décisions — et comment ces décisions sont prises — pour l'examen critique des systèmes d'IA et de leurs répercussions, que ce soit en matière de fiabilité des connaissances produites ou de leur valeur éthique. La documentation des choix effectués durant le développement du modèle d'IA peut faciliter cette supervision (Selbst et Barocas, 2018).

Dans le contexte de l'IA, un biais peut avoir plusieurs significations

Le terme *biais* est souvent utilisé en IA; cependant, il peut avoir différentes significations et être interprété différemment, ce qui peut introduire de la confusion (Hellström *et al.*, 2020). Dans l'usage courant et dans de nombreux comptes rendus médiatiques, le terme est généralement employé pour décrire des modèles d'IA qui produisent des résultats discriminatoires — « quand des jugements injustes sont posés parce que la personne qui les pose est influencée par une caractéristique qui n'est en fait pas rattachée au sujet, généralement une préconception discriminatoire des membres d'un groupe » [traduction libre] (Muller, 2020). Ce type de biais soulève des questions sociales et éthiques, qui sont examinées au chapitre 5. Cependant, il est important de faire la distinction entre ce genre de biais social et éthique et le biais *statistique*, qui peut résulter d'une variété de facteurs, comme des erreurs d'échantillonnage ou de mesure.

Enfin, il peut y avoir un autre *biais* établi en IA, le biais nécessaire pour que le système fonctionne. Par exemple, pour qu'un système d'apprentissage machine puisse faire des inférences inductives afin de généraliser, à partir de ses données d'apprentissage, à de nouveaux exemples, il a besoin d'une certaine forme de biais intégré fondé sur les hypothèses concernant la source des données. C'est ce qu'on appelle le *biais inductif* (Amit et Meir, 2019; Hellström *et al.*, 2020). De même, les concepteurs de systèmes d'IA doivent faire des choix en ce qui concerne le seuil statistique auquel une classification est acceptée (*biais d'incertitude*) et les caractéristiques qui constituent les intrants et les extrants du modèle (*biais de spécification*) (Hellström *et al.*, 2020).

3.1.4 Gérance et gestion des données

Garantir l'intégrité scientifique dans la recherche utilisant l'IA requiert le suivi de principes de gérance et de gestion qui facilitent le partage et utilisation responsables et éthiques des données. Le concept FAIR (Findable, Accessible, Interoperable, Reusable — soit des données faciles à trouver, accessibles, interopérables et réutilisables), qui vise à faciliter la réutilisation des données scientifiques (Wilkinson *et al.*, 2016), est aujourd'hui un principe largement reconnu de gestion et de gérance des données en sciences (Mons *et al.*, 2017; Boeckhout *et al.*, 2018). Constatant l'importance de ce concept pour l'IA en sciences, le Département de l'énergie des États-Unis a affirmé que « les systèmes d'IA peuvent être employés pour automatiser la création de données FAIR et les intégrer dans les dépôts de connaissances, afin de procurer les fondations architecturales de la nouvelle infrastructure de données nécessaire pour accélérer l'apprentissage en IA et la conception de modèles » [traduction libre] (Stevens *et al.*, 2020).

La gérance et la gestion responsables et éthiques des données en sciences et en génie peuvent exiger des mesures dépassant le respect des principes FAIR

Bien que les principes FAIR touchent des questions pratiques concernant le partage et la distribution des données, ils ne portent pas sur les éléments sociaux et éthiques de la gérance et de la gestion des données. Par conséquent, d'autres principes de gestion des données ont été proposés pour les compléter; par exemple :

- Les principes TRUST (Transparency, Responsibility, User focus, Sustainability, Technology — soit la transparence, la responsabilité, la prédominance de l'utilisateur, la durabilité et la technologie), qui concernent la fiabilité des dépôts numériques et visent à fournir un cadre facilitant l'examen et la mise en œuvre de la conservation et de l'archivage des données (Lin *et al.*, 2020).
- Les principes the FACT (Fairness, Accuracy, Confidentiality, Transparency — soit l'équité, l'exactitude, la confidentialité et la transparence), élaborés par la

Responsible Data Initiative dans le but de favoriser l'éthique et la responsabilité en science des données en se concentrant sur les défis scientifiques fondamentaux (van der Aalst *et al.*, 2017).

- Les principes CARE (Collective benefit, Authority to control, Responsibility, Ethics — soit le bénéfice collectif, le pouvoir de contrôle, la responsabilité et l'éthique), qui portent sur la gouvernance des données autochtones et qui ont été élaborés par l'Indigenous Data Sovereignty Interest Group (au sein de la Research Data Alliance (RDA)) en complément aux principes FAIR afin de souligner le rôle des données dans l'amélioration de l'autodétermination et de l'innovation au sein des peuples autochtones (Carroll *et al.*, 2020) (section 5.2.1).

De plus, certains organismes cherchent à régler le problème du partage des données, comme la RDA, un organisme international voué au développement et à la diffusion de l'infrastructure technique, sociale et communautaire nécessaire pour faciliter le partage des données et l'exploration fondée sur les données, particulièrement dans le milieu de la recherche universitaire (Berman, 2019). Au sein de la RDA, des chercheurs de différentes disciplines et différents champs de recherche scientifique mettent sur pied des groupes d'intérêt axés sur les besoins en infrastructure uniques — par exemple en matière de code, de protocoles, d'outils, de modèles, de programmes, de politiques et de normes — de la recherche fondée sur les données dans leur domaine (Berman, 2019).

3.2 Utilisation éthique de l'IA dans la recherche en sciences et en génie

L'utilisation éthique a suscité énormément d'attention, en particulier dans des secteurs tels que la prise de décision (par le gouvernement comme par les organismes privés), l'application de la loi, la reconnaissance faciale, la confidentialité des données en ligne, la désinformation et l'hypertrucage, la manipulation du comportement et les systèmes autonomes tels que les véhicules et les armes (Muller, 2020). On a cependant moins étudié les questions éthiques particulières à l'utilisation de l'IA en sciences et en génie. En fait, bien que les cadres éthiques prolifèrent en ce qui concerne l'IA, la plupart d'entre eux — comme la Déclaration de Montréal IA responsable (2018) — ne portent pas explicitement sur l'utilisation de l'IA dans ces domaines. Il existe cependant une exception, la Recommandation sur l'éthique de l'intelligence artificielle, publiée par l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO, 2020). Celle-ci reconnaît explicitement que l'utilisation de l'IA dans les pratiques des sciences et du génie soulève des préoccupations éthiques fondamentales et encourage la connaissance au sein des communautés scientifiques des avantages, des limites et des risques de cette utilisation, et incite notamment à s'assurer que les conclusions scientifiques

produites sont « solides et fiables » (UNESCO, 2020). De plus, les principes d'Asilomar sur l'IA, élaborés par des experts à la Conférence sur l'IA bénéfique de 2017, sont 23 principes d'utilisation responsable et bénéfique de l'IA concernant la recherche, l'éthique et les valeurs (FLI, 2017a, 2017b).



« Des questions éthiques surgissent à tous les stades de l'utilisation de l'IA dans la recherche scientifique, que ce soit au recueil et au prétraitement des données, à la conception et à l'utilisation des modèles d'IA entraînés avec ces données, à la diffusion et à la publication des résultats ou encore au stockage, à la tenue à jour et à l'accès à long terme aux données, modèles et résultats. »

En dépit du manque généralisé d'orientation éthique concernant l'utilisation de l'IA dans le contexte scientifique et du génie, comme Metcalf *et al.* (2021) le soulignent, les nouvelles formes de production de savoir peuvent exiger de nouveaux cadres éthiques. En fait, dans la mesure où les pratiques existantes d'éthique de la recherche ont été élaborées avant l'actuelle transformation vers des formes algorithmiques de production du savoir (telles que l'IA), on peut s'attendre à des non-concordances entre ces cadres et les exigences éthiques de la recherche guidée par les données (Metcalf *et al.*, 2021). Des questions éthiques surgissent à tous les stades de l'utilisation de l'IA dans la recherche scientifique, que ce soit au recueil et au prétraitement des données, à la conception et à l'utilisation des modèles d'IA entraînés avec ces données, à la diffusion et à la publication des résultats ou encore au stockage, à la tenue à jour et à l'accès à long terme aux données, modèles et résultats.

3.2.1 L'éthique dans le recueil et l'utilisation des données

Veiller au caractère éthique des pratiques de recueil et d'utilisation des données pour l'apprentissage et l'application des modèles est essentiel à l'utilisation responsable de l'IA en sciences et en génie. Le recueil et l'utilisation des données comportent deux exigences particulières : (i) faire en sorte que les ensembles de données ne sont pas faussés ou biaisés de façon que les modèles d'IA entraînés avec eux créent une discrimination contre des individus ou des groupes et (ii) veiller au respect de principes tels que le *consentement éclairé*. La question du biais et de la discrimination dans les ensembles de données est traitée au chapitre 5, alors que la question du consentement éclairé est examinée ci-dessous.

L'IA complexifie les notions traditionnelles de participants humains aux recherches et de consentement éclairé

Un des problèmes éthiques les plus urgents à propos de l'utilisation de l'IA en sciences concerne la recherche reposant sur des participants humains. Ce type de recherche est depuis longtemps assujéti à des exigences éthiques particulières. Au Canada, l'*Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains – EPTC 2*, publié par les IRSC, le CRSNG et le CRSH, dont la dernière version date de 2018, comporte un ensemble largement utilisé d'exigences de ce type. L'EPTC 2 établit les principes et les politiques auxquels les chercheurs et leurs établissements doivent adhérer pour être admissibles au financement. Bien que ce document ne mentionne pas explicitement l'IA, il décrit certains éléments éthiques concernant les mégadonnées, tels que la possibilité de réidentifier des données précédemment dépersonnalisées (IRSC *et al.*, 2018b)⁵.

Fait important, cependant, l'utilisation des mégadonnées et des outils d'IA dans la recherche scientifique complexifie le concept de participants humains aux recherches. Alors que l'on considère généralement que ce terme fait référence aux personnes qui fournissent un consentement éclairé en toute connaissance de cause et volontairement pour participer à une expérience de recherche, la recherche employant les mégadonnées et l'IA peut utiliser des ensembles de données contenant des renseignements sur des personnes sans que celles-ci le sachent ou qu'elles y aient consenti⁶. Ce problème est particulièrement évident dans le cas de l'utilisation de données collectées par le biais des réseaux sociaux (encadré 3.1).

5 Les *données dépersonnalisées* sont des données dans lesquelles les renseignements personnels ont été supprimés (ICPO, 2016). Cependant, il existe des risques de réidentification des renseignements dépersonnalisés, ce qui les distingue des données anonymisées, qui sont des données pour lesquelles la réidentification est impossible (Thompson et Lucarini, 2021). L'Énoncé de politique des trois conseils recommande d'utiliser des données anonymisées lorsque c'est possible, et des données dépersonnalisées comme meilleure solution de remplacement (IRSC *et al.*, 2018b).

6 Si l'EPTC 2 (article 3.7) prévoit certaines exemptions à l'exigence d'obtenir le consentement éclairé avant d'utiliser des données personnelles, ces exemptions doivent être vérifiées et approuvées par un comité d'éthique de la recherche (IRSC *et al.*, 2018b). Toutefois, l'examen n'est pas nécessairement effectué dans tous les contextes de recherche qui recourent aux mégadonnées ou à l'IA.

Encadré 3.1 Participants à la recherche, consentement éclairé et données provenant des réseaux sociaux : étude sur la « contagion émotionnelle » de Facebook

En 2014, des chercheurs de Facebook et de l'Université Cornell ont publié les résultats d'une expérience dans laquelle ils avaient manipulé l'algorithme du fil de nouvelles de Facebook de 689 003 personnes afin d'évaluer l'impact de la valence émotionnelle des statuts (Kramer *et al.*, 2014). Cependant, aucun des individus concernés n'avait directement consenti à participer à l'étude ou même été conscient qu'il participait à une expérience; en fait, l'expérience avait été autorisée en vertu de la politique d'utilisation des données de Facebook (Verma, 2014).

L'étude a été critiquée parce que le consentement éclairé des sujets n'avait pas été obtenu; cependant, tous les éthiciens ne s'accordent pas pour dire que l'étude de Facebook était nécessairement non éthique. Dans une réponse aux critiques de l'étude, plus de 30 éthiciens⁷ ont signé une déclaration affirmant que la recherche n'avait pas violé la vie privée des utilisateurs de Facebook et n'était pas substantiellement différente des modifications courantes apportées par Facebook aux algorithmes de son fil de nouvelles (Meyer, 2014). Les signataires y affirment que le consentement éclairé explicite n'était pas requis parce que la recherche était cohérente avec les attentes raisonnables des utilisateurs de Facebook, elle ne posait que peu ou pas de risques additionnels et obtenir le consentement éclairé aurait biaisé les résultats. Cependant, tous convenaient que l'approbation d'un comité de recherche institutionnelle aurait dû être obtenue (Meyer, 2014). Les signataires y avancent également qu'une critique injustifiée de l'étude pourrait dissuader des recherches précieuses et faire en sorte que des études similaires sont effectuées dans le secret et non publiées.

On défend souvent l'utilisation de données personnelles en l'absence de consentement éclairé explicite en arguant que ces activités n'impliquent pas d'intervention directe dans la vie du sujet ou que les renseignements concernés sont du domaine public (Metcalfe et Crawford, 2016; Metcalfe *et al.*, 2021). Ces justifications reposent sur la présomption que ce qui compte, c'est quel genre de données sont recueillies et *comment* elles sont recueillies et ce *qui est fait* avec les données après leur collecte (Metcalfe et Crawford, 2016). En fait, la facilité relative

7 Par souci de transparence, il faut savoir que l'actuel président-directeur général du CAC, Eric M. Meslin, était l'un des signataires. Il n'était toutefois pas affilié au CAC à l'époque.

et le faible coût de recueillir, de stocker et de réanalyser de gros ensembles de données complexifient notablement le concept traditionnel de consentement éclairé, parce que le consentement qui est obtenu au début d'une recherche sera probablement insuffisamment éclairé pour couvrir adéquatement toutes les utilisations possibles des données personnelles d'un individu et les risques rattachés (Metcalf *et al.*, 2021). Ce problème est partiellement traité dans l'EPTC 2 (article 5.5), qui définit les exigences que les chercheurs doivent satisfaire pour utiliser des renseignements à des fins secondaires sans obtenir un nouveau consentement des participants (IRSC *et al.*, 2018b).

De nombreux comités d'éthique de la recherche considèrent que les données publiques sont exemptées d'un examen institutionnel (Metcalf et Crawford, 2016; Leonelli *et al.*, 2021). Par exemple, l'EPTC 2 (article 2.2) comprend une exemption pour la recherche qui s'appuie exclusivement sur des renseignements du domaine public et pour lesquels les individus n'ont pas d'attentes raisonnables en ce qui concerne la confidentialité, tant qu'il n'y a pas d'interaction en ligne directe entre les chercheurs et les sujets (IRSC *et al.*, 2018b). Cependant, l'hypothèse que l'utilisation de renseignements d'accès public aux fins de la recherche ne pose pas de risque et ne cause pas de préjudice néglige le fait que les mégadonnées et les outils d'IA sont capables de générer des renseignements imprévus ou que les données peuvent être utilisées à des fins que les individus consentants pourraient ne pas avoir raisonnablement prévues. C'est particulièrement vrai quand des données provenant de sources variées sont agrégées (Metcalf et Crawford, 2016; Metcalf *et al.*, 2021).

Les données publiques obtenues à partir des réseaux sociaux sans que les sujets de ces données ne le sachent ou n'y aient consenti sont un exemple qui pourrait soulever des questions éthiques ou causer des préjudices (encadré 3.2). Les réseaux sociaux fournissent aux chercheurs une immense quantité de données à une échelle qui était impossible jusqu'ici, et pourraient procurer des renseignements précieux sur le plan scientifique (et commercial) sur une vaste gamme de sujets (Leonelli *et al.*, 2021). Il existe plusieurs exemples de recherche scientifique ayant utilisé ces données dans des initiatives de « science participative » (Ceccaroni *et al.*, 2019) et dans la recherche sur la santé (Leonelli *et al.*, 2021). Cependant, l'utilisation des données des réseaux sociaux dans la recherche risque également d'exacerber les inégalités et les injustices sociales (Leonelli *et al.*, 2021). De plus, on s'inquiète également que l'utilisation de ces données en recherche puisse produire des résultats scientifiquement douteux ou non fiables sur le plan épistémique, et même provoquer une réaction négative de la population ou nuire à la perception que le public a de la science (Leonelli *et al.*, 2021). Plusieurs tentatives ont d'ailleurs été effectuées pour concevoir des cadres éthiques de l'utilisation des données provenant des réseaux sociaux. De plus, Leonelli *et al.* (2021) définissent les étapes

que les chercheurs peuvent suivre pour mettre en œuvre une *équité des données méthodologiques* — qui se concentre sur « la qualité et la crédibilité des processus de recherche par lesquels les données sont produites, rassemblées, mises en commun, analysées et interprétées » [traduction libre] — pour répondre aux questions épistémiques et éthiques soulevées par l'utilisation des données (y compris des données des réseaux sociaux) à des fins de recherche.

Encadré 3.2 Opposition entre respect de la vie privée et mégadonnées : le cas de Clearview AI

Clearview AI a conçu un système de reconnaissance faciale articulé autour de l'IA qui « moissonne » des photos de visages (et les données connexes) à partir de sources publiques, y compris des réseaux sociaux, sans la connaissance et le consentement des individus. L'entreprise stocke ensuite ses renseignements dans une base de données qui est utilisée par ses clients, tels que les organisations commerciales et les forces de police — comme la Gendarmerie royale du Canada (GRC), à l'époque —, qui peuvent ensuite téléverser des images pour trouver des correspondances dans la base de données (OPC, 2021c; Thompson, 2021). Une enquête conjointe menée par le Commissariat à la protection de la vie privée du Canada et la Commission d'accès à l'information du Québec et les Commissariats à l'information et à la protection de la vie privée de la Colombie-Britannique de l'Alberta a déterminé que Clearview AI avait violé les lois fédérale et provinciales concernant le recueil, l'utilisation et la communication de renseignements personnels⁸. Cependant, Clearview AI a contesté ces conclusions, a refusé de reconnaître avoir commis un méfait et ne s'est pas engagée à suivre les recommandations ou les ordonnances des autorités canadiennes (OPC, 2021c). La capacité des autorités canadiennes à pénaliser l'entreprise ou à la forcer à respecter le droit canadien est limitée parce que Clearview AI est une entreprise basée aux États-Unis (Thompson, 2021).

Bien que cet exemple dépasse la portée du présent rapport parce que les renseignements recueillis par Clearview AI n'étaient pas utilisés pour la conception ou la découverte en sciences ou en génie, il souligne les dangers de la combinaison des renseignements personnels publics et de l'IA et la difficulté de faire appliquer la loi sur le respect de la vie privée d'un pays à l'autre en ce qui concerne mégadonnées.

8 Dans une enquête ultérieure, le Commissariat à la protection de la vie privée du Canada a constaté que la GRC avait violé le droit canadien en recourant de façon illégale aux services de Clearview IA (OPC, 2021b). La GRC a également exprimé son désaccord avec les conclusions du Commissariat, mais elle a convenu de mettre en œuvre ses recommandations (Boutilier, 2021; OPC, 2021b).

En fin de compte, la popularisation des mégadonnées et de l'IA peut nécessiter d'abandonner la notion traditionnelle de consentement éclairé, focalisée sur le recueil des données, au profit d'un consentement à l'utilisation de données qui existent déjà. Ce changement est partiellement évident avec le *Règlement général sur la protection des données* (RGPD) de l'UE, qui porte en grande partie sur le traitement des données personnelles plutôt que sur leur collecte. Fait notable, le RGPD (article 89-2) permet aux États membres de l'UE d'accorder des exemptions



« En fin de compte, la popularisation des mégadonnées et de l'IA peut nécessiter d'abandonner la notion traditionnelle de consentement éclairé, focalisée sur le *recueil* des données, au profit d'un consentement à l'*utilisation* de données qui existent déjà. »

à certaines de ses exigences (comme le droit d'accès, la rectification, la limitation du traitement et l'objection) quand le traitement des données est effectué à des fins scientifiques, tant que des balises appropriées ont été mises en place (Mourby *et al.*, 2019; RGPD, 2021).

Face à l'intensification de la recherche très gourmande en données, de nouveaux modèles de consentement éclairé sont en cours d'élaboration, par exemple pour les données sur la santé (Vayena et Blasimme, 2017). Bien que les questions concernant l'utilisation de l'IA pour la recherche en santé et dans le domaine médical dépassent largement la portée de ce rapport, nombre de ces modèles pourraient être appliqués à la recherche en sciences et en génie en général. Par exemple, le modèle de *consentement dynamique* transforme

le consentement d'une décision statique et ponctuelle en un processus adaptatif dans lequel les participants indiquent quelles données pourront ou ne pourront pas être utilisées dans de futures recherches, par le biais d'une communication permanente avec les chercheurs (Vayena et Blasimme, 2017). En outre, des approches technologiques et des mécanismes de gouvernance des données novateurs peuvent aider à protéger le consentement éclairé, comme la gestion électronique du consentement et les coopératives de données participatives.

L'utilisation éthique de l'IA dans la recherche en sciences et en génie doit se concentrer sur les préjudices sociaux en plus des préjudices sur les individus

Bien des techniques telles que l'anonymisation des données personnelles vise à atténuer les préjudices potentiels sur les individus, « l'agrégation massive des données de recherche éloigne notre concept de sujet humain des individus pour l'amener vers des regroupements ou des classifications distribués » [traduction libre] (Metcalf *et al.*, 2021). En fait, le risque que l'IA cause des préjudices à des

groupes historiquement marginalisés est bien connu (section 5.2). De plus, on a avancé que les notions de consentement éclairé axées seulement sur les individus et non sur les groupes (et parfois également sur les données recueillies et la façon dont elles le sont, tout en ignorant comment elles seront utilisées) peuvent créer de graves difficultés pour les communautés autochtones (section 5.3) et contribuer à leur méfiance générale à l'égard de la recherche scientifique (J.E. Lewis, communication personnelle, 2021).

Malgré le fait que l'utilisation de l'IA dans la recherche en sciences et en génie puisse poser des risques sociaux et éthiques dépassant les préjudices individuels, les comités d'éthique de la recherche et les comités d'examen institutionnels classiques « sont conçus pour évaluer les préjudices aux sujets humains, plutôt que les préjudices à la société humaine » [traduction libre] (Bernstein *et al.*, 2021). Face à ces lacunes, des chercheurs de l'Université Stanford ont introduit le concept de comité d'examen éthique et sociétal, qui vise à atténuer les conséquences sociales et éthiques négatives potentielles de la recherche en IA s'inscrivant comme une exigence de l'accès au financement (Bernstein *et al.*, 2021). Les chercheurs doivent soumettre un énoncé présentant les risques de leur projet pour la société (c.-à-d. la société ciblée par la recherche), les sous-groupes sociaux (en particulier les groupes marginalisés) et la collectivité dans son ensemble. De plus, ils doivent définir des principes pour atténuer ces risques et expliquer comment ces principes seront mis en application dans le concept de recherche. Ces énoncés sont ensuite examinés par un comité interdisciplinaire, afin de déceler d'autres risques et établir d'autres stratégies d'atténuation dans le cadre d'une collaboration itérative avec les chercheurs.

3.2.2 L'éthique et la diffusion des constatations relatives à l'IA

Partnership on IA (2021) a examiné comment diffuser de façon responsable les résultats de la recherche en IA, d'une façon qui tienne compte des conséquences en aval et des répercussions générales des conclusions de recherche produites par l'IA. Le rapport formule plusieurs recommandations pour trois entités principales : les chercheurs, la direction de la recherche et les conférences et revues. En ce qui concerne les chercheurs, il recommande la communication proactive de détails supplémentaires sur les travaux, notamment sur leur contribution et leur motivation, des conséquences en aval potentielles et du volume de ressources de calcul utilisées. La direction de la recherche est encouragée à favoriser l'étude précoce des conséquences en aval dans le cadre de l'examen interne, avant d'entreprendre les travaux, et à féliciter et récompenser les chercheurs qui décèlent ces conséquences dans leurs travaux et dans ceux de leurs collègues. Enfin, Partnership on IA (2021) recommande que les responsables des conférences et des revues élargissent leurs critères d'examen par les pairs

pour y inclure la prise en compte des conséquences en aval et établissent des processus d'examen distincts pour les évaluer. Comme exemple de ce type de pratique, en plus de la communication de la liste de contrôle mentionnée à la section 3.1.1, la conférence sur l'IA NeurIPS exige aussi des chercheurs qu'ils soumettent un énoncé des répercussions générales (NeurIPS, 2020), qui décrit « les impacts positifs et négatifs anticipés de leur article et motive ces impacts avec les références appropriées » [traduction libre] (Hecht *et al.*, 2018). Cependant, comme certains pairs examinateurs et rédacteurs en chef de revues et de conférences ne connaissent pas toujours bien les répercussions aval potentielles de la diffusion de la recherche au moyen de l'IA, il peut être nécessaire de leur offrir de la formation à cet égard. Ce besoin est semblable à la nécessité de formation en éthique des futurs scientifiques et développeurs en IA (section 4.5).

3.3 Les répercussions de l'IA sur les valeurs et sur les pratiques sociales en sciences

L'intensification de l'utilisation de l'IA pourrait modifier l'importance relative des différentes valeurs en sciences ainsi que leur dynamique sociale sous-jacente. Les sciences et le génie sont des disciplines intrinsèquement sociales, façonnées par un ensemble particulier de valeurs, notamment des valeurs épistémiques, sociales et éthiques. Les valeurs peuvent changer et l'utilisation de l'IA en sciences peut introduire de nouvelles valeurs, et des conflits peuvent être exacerbés lorsque ces valeurs sont opérationnalisées dans la pratique scientifique. En outre, il est probable que l'IA influe sur la pratique sociale des sciences de nombreuses façons, comme pour la sélection et la conceptualisation des problèmes, l'établissement de la provenance et de la validité des conclusions et l'incidence sur la dynamique sociale de l'interaction humain-IA.

3.3.1 Conflits entre les différents principes et valeurs dans l'IA

L'utilisation responsable et éthique de la recherche concernant l'IA crée souvent un conflit entre différents principes et valeurs, par exemple entre l'*ouverture* et la *rigueur* (ou la *prudence*) (Leonelli, 2020; Whittlestone et Ovadya, 2020). L'engagement à l'égard de l'ouverture exige que la recherche sur l'IA soit aussi ouverte et accessible que possible; cependant, la prudence requiert que l'accès ouvert et la diffusion publique soient limités lorsqu'il y a un risque de mauvais usage (Whittlestone et Ovadya, 2020). La rigueur scientifique demande la surveillance stricte de la manière dont les données sont interprétées et utilisées, ce qui est difficile à contrôler une fois que les données sont librement accessibles (Leonelli, 2020). Les décisions quant à la façon d'équilibrer différentes valeurs et de résoudre les conflits potentiels dans l'application de ces principes doivent

évaluer le degré de risque de mauvais usage par rapport au manque de transparence, l'efficacité de limiter l'accès afin de prévenir ce mauvais usage et les besoins futurs d'empêcher l'utilisation inappropriée et de promouvoir la transparence (Whittlestone et Ovadya, 2020). Il existe d'autres conflits possibles entre les valeurs, comme celui opposant le respect de la confidentialité et des droits aux données et la qualité des ensembles de données et des résultats de l'IA (Muller, 2020), les valeurs sous-tendant la science ouverte et le respect de la vie privée, de la confidentialité et de la sécurité (Science International, 2015) et l'explicabilité et la transparence, d'une part, et l'exactitude prédictive discutée à la section 3.1.2, d'autre part.

Dans les faits, la pratique de la science est façonnée par des valeurs épistémiques (p. ex. le pouvoir explicatif, la simplicité ou la portée) et des valeurs sociales et éthiques non épistémiques (Douglas, 2015). Les valeurs déterminent également les voies de recherche empruntées par les scientifiques et ce qui constitue des normes suffisantes sur les données probantes dans un contexte particulier (Douglas, 2015). Dans le domaine de la recherche sur l'IA, une analyse exhaustive des 100 articles les plus souvent cités a révélé que les valeurs les plus prévalentes étaient les performances, l'exploitation des travaux antérieurs, la généralisation, l'efficacité, les données probantes quantitatives et la nouveauté. En revanche, les préjugés possibles concernant les valeurs liées aux principes éthiques et aux droits des utilisateurs apparaissaient rarement, voire jamais ou pratiquement jamais (2 %) dans les discussions (Birhane, 2021). De plus, l'analyse a montré que bien que les valeurs dominantes étaient à première vue techniques, la façon dont elles étaient priorisées et opérationnalisées révélait des hypothèses implicites reposant sur des valeurs.

3.3.2 Sociologie des sciences et de l'IA

Les sciences sont une discipline intrinsèquement sociale, et l'intensification du recours à l'IA dans ce domaine pourrait modifier sa dynamique sociale (OCDE, 2018b). L'accroissement de l'usage de l'IA dans la recherche scientifique pourrait nécessiter la révision des pratiques sociales sous-tendant la diffusion des conclusions scientifiques (publications, conférences), l'établissement de leur validité (examen par les pairs, réplique) et la mention de leur provenance (crédit, citations) (King *et al.*, 2018; OCDE, 2018b). La validité des connaissances scientifiques est en partie un phénomène constitué socialement, qui dépend du degré auquel les observations et leur interprétation sont convenues par une communauté d'experts faisant autorité (King *et al.*, 2018). De façon similaire, la validité du savoir scientifique produit par la machine dépendra de l'acceptation de ses observations (c.-à-d. des données d'entrée) et de son interprétation (c.-à-d. du modèle d'IA) par la communauté scientifique dans son ensemble.

L'intensification de l'emploi de l'IA en sciences et en génie peut influencer sur la pratique sociale scientifique de nombreuses façons

Globalement, les sociologues des sciences effectuent peu de recherches sur la relation entre l'IA et les scientifiques humains ou « sur les questions sociologiques et anthropologiques concernant la collaboration future entre scientifiques humains et systèmes d'IA » [traduction libre] (OCDE, 2018b). Néanmoins, la relation entre les scientifiques et l'IA peut influencer sur la pratique des sciences à de nombreux niveaux, que ce soit dans la décision quant aux problèmes et aux secteurs devant faire l'objet de recherches, dans la structuration ou la conceptualisation des problèmes afin qu'ils puissent être analysés par l'IA ou dans l'interprétation des résultats inhabituels (King *et al.*, 2018; OCDE, 2018b).

L'IA pourrait aussi réduire les limitations et les problèmes de la pratique sociale actuelle des sciences — y compris les encouragements pervers concernant la publication et le financement, les biais cognitifs humains, la « fausse rigueur » et l'insularité — en promouvant la collaboration plutôt que la compétition et l'autopromotion (The Turing Institute, 2021). Toutefois, l'IA pourrait aussi exacerber ces problèmes si elle n'est pas mise en œuvre de façon responsable. Par exemple, si son utilisation prend une importance exagérée dans la détermination de l'allocation des ressources et du financement de la recherche, l'IA pourrait involontairement fermer la porte à des domaines de recherche prometteurs qui n'y ont pas recours (section 4.1).

L'IA pourrait aussi transformer l'exercice de la recherche scientifique au niveau épistémologique en imposant ce que Lowrie (2017) appelle la *rationalité algorithmique*, principe selon lequel l'étude se concentre sur la faisabilité, l'efficacité et l'utilité d'un algorithme ou d'un modèle précis lorsqu'il est appliqué à un domaine, un problème, un type de données ou un ensemble de données particulier. Dans cette approche, les processus et les outils fondés sur l'IA avec lesquels la recherche est réalisée deviennent l'élément prépondérant de l'étude, façonnant ainsi les objectifs et les résultats de cette recherche. De plus, il sera important de comprendre comment l'IA peut contribuer au savoir scientifique et en quoi ce savoir pourrait être différent du savoir scientifique classique produit par l'humain (King *et al.*, 2018). Cette dernière question est particulièrement intéressante pour déterminer si le savoir scientifique issu de la machine peut être ou pourrait devenir *opaque du point de vue épistémique* — c'est-à-dire, inintelligible pour les êtres humains (Nickles, 2018) (section 3.1.2).

Les partenariats entre les humains et l'IA en sciences ont le potentiel de surpasser chaque forme d'intelligence prise séparément, à peu près de la même façon que les équipes combinant humains et IA surpassent généralement les humains et l'IA seuls aux échecs (OCDE, 2018b; The Turing Institute, 2021). La collaboration humain-IA fructueuse en sciences exigera donc la reconnaissance et la compréhension des capacités et des limites respectives de chaque partie.

Conséquences de l'IA sur le système de recherche canadien

- 4.1 L'utilisation de l'IA dans l'allocation de ressources recherche
- 4.2 Adaptation des systèmes d'examen par les pairs à l'IA
- 4.3 Mesure de l'impact de la recherche
- 4.4 Intégrité de la recherche et gouvernance de la conduite de la recherche
- 4.5 Formation et acquisition des compétences

Constatations du chapitre

- L'utilisation de l'IA dans le processus de recherche brouille les frontières entre les sciences naturelles, le génie, les sciences de la santé, les sciences sociales et les sciences humaines. Les demandes de financement et les programmes de recherche traditionnels portant sur un seul domaine risquent d'être moins pertinents parce que l'expertise acquise dans un domaine ne permet pas d'évaluer adéquatement la nature interdisciplinaire de la science guidée par l'IA.
- La capacité de l'IA à prédire l'impact de la recherche scientifique peut être utile aux gouvernements et aux organismes de financement lorsqu'ils doivent prendre des décisions sur l'allocation des ressources et sur les possibilités de financement destinées à la communauté scientifique.
- L'IA peut soutenir le processus d'examen par les pairs utilisé pour évaluer les applications scientifiques prospectives; cependant, son emploi devra faire l'objet d'une vérification attentive pour s'assurer qu'il n'entraîne pas des conséquences négatives involontaires.
- L'utilisation des données ouvertes dans la recherche gagne en popularité, mais les incitations au partage ne sont pas universelles. De nouvelles politiques et de nouveaux investissements visent surmonter les obstacles et à harmoniser la situation entourant les données.
- L'IA redéfinit l'intégrité de la recherche et la gouvernance de la conduite de la recherche au sein de la communauté scientifique.
- La formation des futurs chercheurs de l'ère de l'IA ne se limitera pas à élargir leurs connaissances et leurs compétences techniques. Les universitaires admettent la nécessité d'enseigner aux futurs scientifiques comment penser les dilemmes éthiques découlant du développement et de l'utilisation de l'IA.

Au Canada, l'allocation des ressources publiques aux fins de recherche non gouvernementale est principalement supervisée par trois organismes : les IRSC, le CRSNG et le CRSH (c.-à-d. les trois organismes), qui se répartissent les responsabilités selon trois axes généraux, soit la santé, les sciences naturelles et le génie, et les sciences sociales et humaines, respectivement. Le système de financement de la recherche universitaire et sa supervision comportent plusieurs étapes de prise de décision concernant les chercheurs, les établissements et les organismes subventionnaires. Les organismes subventionnaires rendent le financement disponible par le biais d'appels à proposition lancés à la communauté

scientifique. Les chercheurs soumettent ensuite leur candidature sous la forme d'une proposition qu'ils ont préparée. Les organismes subventionnaires attribuent généralement les fonds au mérite, lequel est déterminé par un examen par les pairs. Enfin, la gestion après l'attribution implique une prise de décision par l'organisme subventionnaire, les chercheurs bénéficiaires et leur établissement. En tant que technologie décisionnelle, l'IA appliquée à n'importe laquelle de ces étapes aura des conséquences importantes sur le système de financement de la recherche et sur sa gouvernance.

4.1 L'utilisation de l'IA dans l'allocation de ressources recherche

L'utilisation de l'IA dans le processus de recherche brouille les frontières entre les sciences naturelles, le génie, les sciences de la santé, les sciences sociales et les sciences humaines, ce qui remet en question l'utilité des concours de financement et des programmes de recherche traditionnels portant sur une seule discipline. Par conséquent, l'IA amplifie l'observation faite par le Conseil consultatif pour l'examen du soutien fédéral à la recherche fondamentale qu'il est nécessaire de renforcer l'appui à la recherche multidisciplinaire (Naylor *et al.*, 2017). Les sciences humaines et les sciences sociales auront un important rôle à jouer dans la R-D en sciences et en génie qui a recours à l'IA. Il se peut que l'on ait à repenser les divisions disciplinaires traditionnelles du financement de la recherche pour assurer une évaluation équitable et appropriée de la recherche qui emploie l'IA.

Le financement de la recherche est un processus complexe d'équilibrage des intérêts politiques, économiques et scientifiques afin de déterminer si l'attribution de fonds est appropriée et dans quels domaines. Par exemple, accorder la priorité au transfert de connaissances et à la recherche appliquée — qui sont vus comme offrant des possibilités de rendement commercial immédiat — par rapport à la recherche de base (c.-à-d. stimulée par la curiosité) a entraîné la restructuration de l'écosystème de recherche du Canada (Naylor *et al.*, 2017). On attend de plus en plus des organismes de financement qu'ils produisent des offres de financement à une population croissante de chercheurs avec des ressources comparativement moindres (CAC, 2021a). Les efforts visant à corriger cette situation en refondant les structures d'aide aux décisions de financement peuvent créer d'autres problèmes, comme le sous-financement de certains chercheurs en raison du stade où ils en sont dans leur carrière ou de la nouveauté de leur domaine (Reardon, 2015) ou de la baisse de qualité des examens en raison de la virtualisation de l'examen par les pairs (Webster, 2015; Woodgett, 2018). Maintenant que les systèmes d'IA ont été adoptés par les éditeurs et les organisateurs de conférences (Heaven, 2018; Hutson, 2021), il est raisonnable de penser que ces systèmes seront intégrés aux mécanismes d'allocation des ressources dans l'avenir.

En prédisant l'impact de la recherche, les systèmes d'IA peuvent de plus en plus déterminer l'évolution de la découverte scientifique

La capacité de l'IA à prédire l'impact de la recherche scientifique (section 4.3) peut être utile aux gouvernements et aux organismes de financement lorsqu'ils doivent décider s'il faut allouer des ressources à la communauté scientifique, et quand les allouer. L'IA peut également guider et déterminer les priorités de recherche de ces organismes. Comme Weis et Jacobsen (2021) le suggèrent, un programme d'IA (p. ex. le programme Dynamic Early-warning by Learning to Predict High Impact) pourrait être utilisé pour « aider à élaborer des stratégies de financement [en comblant les lacunes sur le terrain de la recherche par] des nouvelles possibilités de programmes de recherche conçus pour optimiser les relations de fort impact prédit » [traduction libre]. Cet emploi de l'IA soulève des préoccupations parce qu'il peut aboutir à une prophétie autoréalisatrice : si plus d'argent est alloué à un certain domaine parce qu'on a prédit qu'il aurait un fort impact, davantage de chercheurs travailleront dans ce domaine et produiront donc vraisemblablement plus de résultats qui seront comptabilisés comme un « impact » (Chawla, 2021). Cela ne signifie pas nécessairement que l'allocation du financement servira mieux les systèmes de recherche ou la découverte canadienne dans son ensemble.

Comme l'IA joue un rôle de plus en plus central dans le cycle de la découverte, le processus de demande de financement de la part des chercheurs devrait changer.

L'attribution de subventions de recherche appuie les cycles de découverte scientifique — l'emploi des connaissances préalables pour formuler des hypothèses, la vérification des hypothèses au moyen de l'expérimentation, l'interprétation et l'analyse des résultats expérimentaux et la contribution des théories définitives aux connaissances préalables. L'intégrité de ce cycle peut être compromise à chaque étape par les limitations et les lacunes propres à l'humain, comme l'incomplétude, l'erreur et le biais. La compétitivité croissante du système de financement de la recherche peut avoir comme effet contre-intuitif d'induire un tel comportement (The Turing Institute, 2021). Comme les limites humaines sont de plus en plus apparentes dans la gestion des forts volumes de données demandant un traitement rapide et exact, les besoins de solutions augmentent, à tel point qu'il est plausible que les fossés dans lesquels ces limitations se produisent soient comblés par un « système d'IA entièrement automatisé » (King *et al.*, 2009; Sparkes *et al.*, 2010; The Turing Institute, 2021). Ces changements transformeront probablement le processus de demande de financement lui-même.

L'IA possède le potentiel de stimuler la recherche scientifique en permettant la génération d'hypothèses, la conception d'expériences, l'expérimentation, l'interprétation et l'analyse automatiques

Les propositions de subvention de la recherche proviennent généralement des chercheurs, et non d'une vision stratégique ou d'un commanditaire; c'est-à-dire que les hypothèses qu'on se propose d'étudier découlent de la compréhension que le chercheur a des connaissances préalables dans son domaine. Or, en raison de l'expansion continue des connaissances préalables, le simple fait de se tenir au courant dans son propre champ d'expertise est devenu complexe, une situation qui nécessitera « un examen automatisé des articles » (The Turing Institute, 2021). L'IA est développée pour fournir aux examinateurs un résumé général des articles, un sommaire des contributions, une analyse de la qualité d'écriture et les travaux

connexes afin de réduire le temps nécessaire à l'examen par les pairs (Roberts et Fisher, 2020).

Appliquée aux sciences, elle rend également possibles la génération d'hypothèses, la conception d'expérience, l'expérimentation, l'interprétation et l'analyse automatiques, et est déjà utilisée dans tous ces domaines à un certain degré. Par exemple, elle est utilisée pour la lecture, l'interprétation et l'inférence à partir du génome humain afin de prédire quels composés sont des candidats prometteurs au traitement des mutations qui causent la maladie (Wainberg *et al.*, 2018). Deep Genomics, une entreprise canadienne qui tire parti de l'IA pour rechercher de nouveaux médicaments potentiels, suggère que « le futur de la conception des médicaments repose sur l'IA, parce que la biologie est trop complexe pour être comprise par les humains »

[traduction libre] (Deep Genomics, 2021). Dans cette utilisation, l'IA employée dans la découverte scientifique est davantage qu'une autre technologie visant à éliminer la pénibilité des tâches répétitives, comme l'ont été les calculatrices et les ordinateurs. À la place, elle servira plus ou moins à diriger les recherches.

La compétition pour le financement pourrait devenir moins une question de mérite et plus une question d'accès aux ressources d'IA. Imaginons, par exemple, l'effet d'un bras robotisé capable d'effectuer des tests toxicologiques à haut risque en une journée, alors que ces tests auraient pris un an à un humain. L'accès à une telle



« L'IA employée dans la découverte scientifique est davantage qu'une autre technologie visant à éliminer la pénibilité des tâches répétitives, comme l'ont été les calculatrices et les ordinateurs. À la place, elle servira plus ou moins à diriger les recherches. »

technologie a été décrit comme un événement transformateur pour notre compréhension de la toxicologie, en raison de la rapidité des tests (Bogue, 2012). On présume que les demandes de financement déposées par des toxicologues qui disposent d'un bras robotisé dans leur laboratoire deviendront des propositions plus solides et plus complètes que les propositions de ceux qui n'en disposent pas. L'IA peut porter ce type d'avantage un autre niveau en dépassant l'automatisation des laboratoires pour rendre possible la « découverte scientifique robotisée » (Bogue, 2012) ou, comme Sparkes *et al.* (2010) l'appellent « le scientifique robot ». Si la biologie devient trop complexe, les modèles d'IA capables de faire des découvertes novatrices et de réaliser des prédictions exactes offriront aux chercheurs en biologie qui y ont accès un avantage dans le financement concurrentiel.

Une conséquence de ce scénario serait la concentration du financement sur les chercheurs et les organismes avantagés du point de vue économique qui ont accès à l'IA. Bien entendu, bien avant l'utilisation de l'IA dans la découverte, le système de financement de la recherche était déjà guidé par des facteurs autres que la qualité des propositions de recherche (Laudel, 2006). Comme Robert Merton l'affirmait il y a 50 ans, les processus sociaux d'allocation des ressources en sciences créaient déjà l'effet Matthew, en vertu duquel les succès précédents en matière de financement conduisent à l'augmentation du financement (Merton, 1968). L'utilisation de l'IA dans la découverte risque d'introduire une nouvelle sorte d'effet Matthew, un effet non seulement induit par les ressources, mais également par l'intelligence. Dans ce contexte, l'idée de « mérite » — un principe fondamental régissant le financement compétitif de la recherche — sera une qualité de plus en plus définie par une aptitude non humaine.

4.2 Adaptation des systèmes d'examen par les pairs à l'IA

Alors que la concurrence pour le financement s'intensifie, l'examen par les pairs est soumis à une pression croissante, ce qui incite les organismes de financement à apporter des changements à leur processus pour réduire le temps, les ressources et les efforts déployés. Les éditeurs et les organismes de financement utilisent l'automatisation pour vérifier les éventuels plagiats dans les soumissions, pour la conformité au formatage et pour assigner des examinateurs (Charlin et Zemel, 2013; Grant, 2017; Checco *et al.*, 2021). À ce sujet, l'IA sert d'outil de maintien de la qualité scientifique aux premières étapes de l'examen.

L'IA peut soutenir le processus d'examen par les pairs, mais les problèmes de conflits d'intérêts et à de complexité exigeront une attention particulière

L'affectation des examinateurs est une tâche répétitive et extrêmement détaillée, qui se prête aux solutions d'IA. Les éditeurs commencent à filtrer les soumissions à la recherche de conflits d'intérêts entre les chercheurs, les rédacteurs et les examinateurs (Chawla, 2020). Cependant, bien que l'IA permette une meilleure détection des relations entre les scientifiques, il y a des nuances dans les conflits d'intérêts qu'elle n'est pas toujours en mesure de déceler. Ces conflits peuvent être subjectifs et dépendre de la taille du domaine et de la nature des précédents contacts entre les scientifiques (Chawla, 2020). Par exemple, bien qu'un candidat ne devrait pas être évalué par un examinateur du même établissement, ce fonctionnement peut être autorisé lorsqu'il est adéquatement justifié. La détection des conflits ne s'effectue pas de façon binaire : il y a conflit ou pas de conflit; elle s'effectue en fonction de l'évaluation de leur gravité (Lo et Field, 2009). Il est nécessaire de juger « la probabilité que des décisions professionnelles prises dans des circonstances pertinentes soient indûment influencées par un intérêt secondaire [et] la gravité du préjudice ou de l'injustice que pourrait produire une telle influence » [traduction libre] (Lo et Field, 2009).

Lors de l'examen par les pairs, les décisions guidées par des systèmes automatisés peuvent également être obscurcies par la propre opacité de l'algorithme (Checco *et al.*, 2021). Par conséquent, au lieu de résoudre le problème de boîte noire dont souffre l'examen par les pairs (Oransky et Marcus, 2017), l'IA peut simplement remplacer une boîte noire par une autre. Qu'ils soient guidés ou non par l'IA, les organismes de financement doivent faire en sorte que leurs décisions soient interprétables par les demandeurs et par le public si l'on veut que leurs décisions soient transparentes et, en fin de compte, qu'elles suscitent la confiance. Pour certaines demandes, il est tout à fait possible de bâtir des modèles d'apprentissage machine transparents et interprétables par les humains qui fonctionnent de façon aussi exacte que les modèles boîte noire (Rudin, 2019; Rudin et Radin, 2019).

L'IA devra être programmée de manière à favoriser un système d'examen par les pairs équitable, diversifié et inclusif

L'équité, la diversité et l'inclusion (EDI) dans le système de recherche sont une priorité des trois organismes. Un des aboutissements prévus du Plan d'action des trois organismes pour l'EDI (2018–2025) est la diversité des membres dans les comités d'examen (IRSC *et al.*, 2018a). Pour y parvenir, il sera nécessaire de mettre sur pied des politiques et des processus faisant en sorte que l'examen par les pairs soit inclusif et représentatif de la diversité canadienne (IRSC *et al.*, 2018a). L'IA employée pour éclairer la sélection des examinateurs devra être en mesure de

prendre des décisions nuancées fondées sur l'expertise et la diversité. On s'inquiète toutefois que l'IA puisse exacerber la marginalisation des groupes généralement sous-représentés dans le système de recherche (chapitre 5). L'intégration des systèmes et de la technologie d'IA dans le système d'examen par les pairs pourrait atténuer ce problème, mais on ne peut présumer de rien. Elle devra être effectuée avec soin et prendre en considération les limitations de l'IA.

L'utilisation de l'IA dans les examens par les pairs peut faire plus que simplifier des tâches administratives qui prennent du temps et sont propices à l'erreur humaine. Elle peut également servir à déterminer la lisibilité, la pertinence et le respect des exigences de formatage des soumissions, empêchant ainsi les examinateurs de prendre des décisions basées sur des biais de première impression plutôt que sur le mérite scientifique (Checco *et al.*, 2021). Toutefois, l'IA présente ses propres biais. Comme Checco *et al.* (2021) le soulignent, elle est entraînée avec des données passées et est donc intrinsèquement conservatrice. Il existe un risque que les problèmes signalés par le filtrage automatique des soumissions nuisent au rôle de l'examinateur, perpétuant les valeurs et les attentes implicites dans les données à partir desquelles le système a appris (Checco *et al.*, 2021). Cela peut désavantager des groupes sous-représentés dans les publications.

Il est peu probable que remplacer l'humain par un système fondé sur l'IA pour la prise de décision dans le cadre de l'examen par les pairs élimine les biais inhérents ou implicites. Étant donné que de nombreux systèmes d'IA reposent sur des



« Il est peu probable que remplacer l'humain par un système fondé sur l'IA pour la prise de décision dans le cadre de l'examen par les pairs élimine les biais inhérents ou implicites. »

techniques d'apprentissage machine qui extraient les modèles à partir des données utilisées, l'intégrité des ensembles de données est cruciale pour la qualité du système (Muller, 2020). Autrement dit, un système d'IA qui travaille à partir d'un ensemble de données biaisé se transformera en un système biaisé. Mais le biais technique n'est qu'une forme de biais potentiel dans les systèmes d'IA utilisés pour l'examen par les pairs. Les systèmes d'IA « sont façonnés par l'environnement dans lequel ils sont construits et par les gens qui les construisent » [traduction libre] (West *et al.*, 2019). Une « crise de la diversité » dans les espaces sociaux dans lesquels les systèmes d'IA sont bâtis peut causer le traitement

différentiel de plus de gens à mesure que ces systèmes jouent un rôle plus important dans un plus grand nombre de domaines (West *et al.*, 2019). À mesure que l'EDI sera priorisée dans l'examen par les pairs, les systèmes d'IA déployés à cette fin devront être basés sur des ensembles de données et des principes de conception équitables, diversifiés et inclusifs.

4.3 Mesure de l'impact de la recherche

L'IA peut être capable de prédire l'impact de la recherche scientifique mieux que les humains ou que les indicateurs actuellement employés, comme le nombre de citations (c.-à-d. en prévoyant le nombre d'études futures qui feront référence au projet de recherche financé). De la même façon que la capacité prédictive de l'IA pourrait indiquer des domaines d'études prometteurs (Wang et Barabási, 2021; Weis et Jacobson, 2021), les organismes de financement peuvent utiliser cette technologie pour éclairer les décisions concernant le renouvellement des subventions. Bien entendu, les répercussions du financement de la recherche ne se font sentir que des années après que la décision de financement a été prise. Mais les tentatives pour prédéterminer cet impact à l'aide de facteurs autres que les travaux scientifiques eux-mêmes risquent de mener à des conclusions erronées. Par exemple, utiliser les indices de citations d'une revue pour évaluer la recherche présume que la qualité de la revue est représentative de la qualité de l'article — une hypothèse incorrecte, étant donné la diversité des façons de manipuler l'indice de citations d'une revue (Seglen, 1997).

Après avoir « étudié un vaste éventail de fonctions et utilisé seulement celles qui contiennent un réel signal quant au futur impact » [traduction libre], Weis et Jacobson (2021) prétendent que l'IA plus perfectionnée réduit les biais systémiques latents. Néanmoins, il y a un risque que l'IA utilisée pour prédire l'impact de la recherche emploie simplement « des versions toujours plus perfectionnées d'indicateurs fondamentalement inutiles » [traduction libre] (Seglen, 1997). Quelle que soit l'utilité de l'indicateur, la capacité prédictive de l'IA dépend généralement de données historiques. Par conséquent, le biais défavorable aux personnes sans profil de recherche (p. ex. les chercheurs en début de carrière) est peut-être inévitable lorsque l'on s'appuie sur des systèmes d'IA pour guider les décisions de financement à partir de l'impact de la recherche. Comme le déclare Sydney Brenner (1995), à moins que la technologie puisse connaître et lire le contenu scientifique de l'article, l'IA ne peut pas remplacer l'examen d'un expert (Seglen, 1997). Ainsi que le reconnaît la Déclaration de San Francisco sur l'évaluation de la recherche, la recherche doit être évaluée sur ses propres mérites (DORA, s.d.).

De nouvelles politiques et de nouveaux investissements encourageant le partage des données peuvent aider à réaliser le potentiel de l'IA dans le processus de découverte

Les applications d'IA en sciences et en génie exigeront fréquemment un accès aux données qui ne correspond pas nécessairement à la portée des initiatives relatives aux données ouvertes lancées par les organismes gouvernementaux. Elles reposeront plutôt sur la disponibilité des données découlant de la recherche universitaire. Le développement et la croissance de l'IA pour les sciences et le génie dépendent de la

capacité des chercheurs à accéder aux données de recherche et à les manipuler et les stocker, ce qui entraîne des exigences pour les infrastructures de stockage et de partage des données et la définition ou l'amélioration des normes concernant les données et des cadres de gouvernance (The Royal Society, 2019).

La promotion de l'adhésion aux pratiques de données ouvertes en dehors du secteur public se heurte à plusieurs difficultés majeures. Les organismes de financement et les revues scientifiques de premier plan ont incité ou obligé l'usage des données ouvertes à divers degrés (cOAlition S, 2021; Yeston, 2021), mais les encouragements professionnels pour que les chercheurs divulguent les données



« Le développement et la croissance de l'IA pour les sciences et le génie dépendent de la capacité des chercheurs à accéder aux données de recherche et à les manipuler et les stocker, ce qui entraîne des exigences pour les infrastructures de stockage et de partage des données et la définition ou l'amélioration des normes concernant les données et des cadres de gouvernance. »

ont flous quand on les mesure au risque que ces données soient mal utilisées ou que leur paternité ne soit pas créditée (TMS, 2017). Bien que le degré de respect des principes de données ouvertes par les chercheurs canadiens soit difficile à préciser, il n'est pas uniforme selon la discipline et est inférieur à celui de plusieurs autres pays (Larivière et Sugimoto, 2018), malgré les politiques obligeant l'accès ouvert mises en place par les principaux organismes de financement public (IRSC, 2019). De plus, les politiques sur l'accès ouvert des trois organismes portent sur les publications plutôt que sur les données — par exemple, le CRSNG n'exige pas que les données ouvertes respectent sa politique (CRSNG, 2014) — de façon telle que la disponibilité des données dépend de la politique des revues dans lesquelles les chercheurs publient. Une modification apportée en 2021 par les trois organismes à leur politique sur la gestion des données de recherche (GDR) énonce que le principe des données ouvertes n'est pas obligatoire, bien qu'il soit demandé aux chercheurs d'indiquer dans

le plan de GDR qui accompagne leurs demandes de financement les raisons (p. ex. éthiques, légaux ou commerciales) les empêchant de partager leurs données (GC, 2021d). Les bénéficiaires de subvention seront bientôt tenus de stocker le code, les données et les métadonnées qui sous-tendent leurs publications dans un dépôt numérique, mais les chercheurs disposeront d'une certaine liberté dans la définition des données de recherche qu'ils considèrent comme concernées. Tous les établissements qui gèrent des fonds de subvention de la recherche devront mettre sur pied une stratégie de GDR (GC, 2021d).

Les normes et les pratiques concernant la production et l'utilisation des données — en ce qui concerne par exemple le volume, le format et les flux de tâches — diffèrent d'une discipline et d'un établissement à l'autre. De plus, le grand nombre de parties prenantes, dont les priorités et les besoins sont divers, complique la modernisation de la GDR au Canada (Baker *et al.*, 2019). Les dépôts sont des éléments essentiels de l'infrastructure à cet égard, mais ils doivent surmonter plusieurs défis, qui concernent notamment l'harmonisation des normes, l'offre de la capacité de stockage et la création de communautés de pratique. Les dépôts existants ont tendance à fonctionner de façon indépendante et sont structurés selon les besoins des communautés qui les utilisent le plus. Les encouragements à introduire une plus grande interopérabilité peuvent être insuffisants (Baker *et al.*, 2019). Leur financement n'est également pas uniforme⁹ et s'effectue fréquemment selon un processus concurrentiel, ce qui lie leur existence à des projets qui, pour être financés, doivent respecter les politiques de l'organisme subventionnaire. Ce fonctionnement peut saper la résilience parce que les dépôts peuvent être créés et tenus à jour par des chercheurs individuels et donc, être sujets à des lacunes de financement (Baker *et al.*, 2019).

Plusieurs initiatives concernant les données de recherche ont été lancées récemment, qui ont des répercussions sur l'amélioration de l'accès aux données dans l'écosystème de recherche canadien. En 2019, l'Alliance de recherche numérique du Canada (anciennement la Nouvelle organisation d'infrastructure de recherche numérique) a été créée pour mener l'établissement d'un environnement numérique intégré pour la recherche canadienne, et se concentre sur le calcul, la GDR, les logiciels de recherche et la cybersécurité (NDRIO, 2021a; L'Alliance, 2021). L'Alliance a tenu une consultation avec des chercheurs pour confirmer la diversité des besoins et des priorités entre les disciplines, mais le principal besoin global exprimé concernait les dépôts de données (NDRIO, 2021b). Conjointement avec Calcul Canada, elle travaille à la mise sur pied du Dépôt fédéré de données de recherche — une plateforme unique de dépôts de données partagées entre plusieurs établissements — mais cette initiative n'est financée que depuis peu (FRDR, 2021).

Selon Baker *et al.* (2019), le Canada est à la traîne de ses pairs internationaux en matière de ressources humaines et financières consacrées à la GDR et il a manqué l'occasion de participer aux discussions internationales concernant les normes sur les données¹⁰, et ainsi de favoriser un écosystème d'infrastructure de recherche numérique qui appuie les activités nationales de RD sur l'IA. Les vastes dépôts ouverts de données sont essentiels pour faire progresser l'IA pour la conception

9 Par exemple, le stockage des données à long terme n'est pas admissible à l'aide de la Fondation canadienne pour l'innovation, bien que l'environnement de recherche canadien manque de stockage de ce type (Baker *et al.*, 2019).

10 Des efforts nationaux de normalisation sont également en cours (CSC, 2020) et le budget fédéral 2021 prévoit du financement supplémentaire (GC, 2021c).

et la découverte dans de nombreuses disciplines (U.S. NRC, 2014; De Luna *et al.*, 2017; TMS, 2017). Certains pays ont déjà établi des dépôts de cette taille, répartis sur de multiples éléments d'infrastructure régionaux; par exemple, l'Australie a fondé en 2018 l'Australian Research Data Commons afin de regrouper des organismes d'infrastructure de données autrefois distincts, dans le cadre de la National Research Infrastructure Roadmap 2016 (ARDC, 2021). Plus récemment, l'European Open Science Cloud (EOSC) a lancé une plateforme destinée aux données et aux publications relatives aux initiatives de recherche menées dans l'UE conformes aux principes FAIR (EOSC, 2021). Visant plus loin, dans les documents stratégiques qu'il a publiés, l'EOSC nomme directement les ordinateurs parmi les principaux utilisateurs de son infrastructure (EOSC, 2021).

4.4 Intégrité de la recherche et gouvernance de la conduite de la recherche

L'IA transforme la responsabilité au sein de la communauté scientifique, remettant en question les politiques actuelles de conduite de la recherche

Le financement de la recherche est régi en partie par des codes éthiques auxquels les chercheurs et leur établissement doivent se conformer. Le Cadre de référence des trois organismes sur la conduite responsable (CRCR) de la recherche impose aux chercheurs financés par les organismes des responsabilités en ce qui a trait au renforcement de l'intégrité de la recherche et à la lutte contre les mauvais comportements. Les chercheurs doivent fournir des renseignements précis et complets aux organismes lorsqu'ils déposent une demande de financement (IRSC *et al.*, 2016). Le gouvernement du Canada a établi un modèle de politique sur l'intégrité scientifique, qui aide les ministères et les organismes à faire en sorte que leurs recherches intra-muros respectent les divers principes et règles éthiques (GC, 2021e). Également, comme il est dit au chapitre 3, la conduite des chercheurs est également gouvernée par des règles régissant spécifiquement la recherche impliquant des sujets humains. Étant donné que l'IA promet d'accroître l'exactitude et la cohérence, l'exigence de responsabilité des chercheurs devrait changer. Si les systèmes d'IA deviennent la norme pour l'exécution des tâches scientifiques, les attentes concernant la responsabilité en matière d'exactitude et d'exhaustivité pourraient éliminer les travaux réalisés par les humains de toute considération. Cela pourrait empêcher les demandeurs qui n'ont pas accès à l'IA d'être financés pour des raisons d'irresponsabilité présumée.

L'introduction de l'IA dans le processus de recherche complique la notion de non-respect des dispositions du CRCR. Les points examinés au chapitre 3, comme la reproductibilité, l'explicabilité et l'exactitude, sont particulièrement importants ici. Qui, ou quoi, est exactement responsable des non-respects de la politique

contre la fabrication, la falsification, le plagiat et ainsi de suite, quand il est possible de remonter aux tâches effectuées par l'IA? Bien que l'IA soit utilisée pour faire respecter l'intégrité scientifique en détectant les violations des règles de recherche responsable, il se peut également qu'elle permette à des chercheurs



« Qui, ou quoi, est exactement responsable des non-respects de la politique contre la fabrication, la falsification, le plagiat et ainsi de suite, quand il est possible de remonter aux tâches effectuées par l'IA? »

d'avoir un comportement irresponsable. Il pourrait aussi être possible à un algorithme d'IA lui-même de causer de telles violations. Il sera donc important d'établir un modèle « d'IA responsable [qui procure des] mécanismes permettant aux systèmes d'IA d'agir conformément à l'éthique et aux valeurs humaines » [traduction libre] (Dignum, 2020b) pour maintenir l'intégrité de la recherche. Les systèmes d'IA utilisés dans la recherche devront être conçus de façon à exécuter leurs tâches d'une manière conforme à l'éthique de la recherche régissant les processus de financement concurrentiels. Concevoir une IA responsable exigera de tenir explicitement et systématiquement compte des principes de reddition des comptes, de responsabilité et

de transparence (Dignum, 2020b). Cela nécessitera des chercheurs et des développeurs « formés pour être conscients de leur propre responsabilité quand il s'agit de concevoir des systèmes d'IA ayant des répercussions directes sur la société » [traduction libre] (Dignum, 2020b).

Par exemple, la recherche touchant les communautés autochtones (section 5.3) impose des responsabilités aux chercheurs et à leur modèle d'IA qui rejoignent les préoccupations plus vastes concernant les relations entre l'IA et les sujets de recherche humains (section 3.2). La participation de la population est particulièrement importante pour toute recherche ayant une incidence sur les communautés autochtones (IRSC *et al.*, 2018b). Étant donné que les sciences et les technologies ont été longtemps utilisées d'une façon qui n'a pas profité aux peuples et aux communautés autochtones (IRSC *et al.*, 2018b), l'utilisation de l'IA dans la recherche concernant ces peuples peut se heurter à de la suspicion. Pour être fructueuse, l'IA devra être harmonisée avec le *protocole autochtone* — les coutumes, les traditions et les codes qui régissent le comportement — et trouver sa place au sein des cercles de relations existants (Lewis, 2020)¹¹. L'utilisation de l'IA dans la recherche concernant les peuples autochtones devra donc être comprise par les personnes touchées, de façon que les présuppositions culturelles encodées dans la technologie soient admises et puissent être réorganisées (Lewis, 2020).

11 Sur l'importance d'accepter les traditions autochtones pour acquérir une plus grande compréhension des conséquences de la pensée occidentale, voir Borrows (2010, 2012).

4.5 Formation et acquisition des compétences

Le processus de découverte de demain nécessitera l'acquisition d'expertises en matière d'IA

Avec l'expansion de l'utilisation de l'IA, davantage de domaines vont modifier leurs attentes en ce qui concerne les compétences et les connaissances des chercheurs (Fleming, 2018). Les programmes d'études en sciences changeront également pour tenir compte des transformations dans les compétences exigées par l'IA. Il faudra probablement avoir recours à l'acquisition des compétences et à la formation pour permettre à la main-d'œuvre de s'adapter aux changements découlant de l'utilisation de l'IA dans les laboratoires (Lane et Saint-Martin, 2021). Par exemple, il se peut que les chercheurs aient besoin de formation pour apprendre à utiliser efficacement les outils de science des données (Ezer et



« La formation des futurs chercheurs de l'ère de l'IA ne se limitera pas à élargir leurs connaissances et leurs compétences techniques. Les universitaires admettent la nécessité d'enseigner aux futurs scientifiques comment penser les dilemmes éthiques, culturels et sociaux découlant du développement de l'utilisation de l'IA. »

Whitaker, 2019). Selon une analyse de l'OCDE (2021b) réalisée au moyen de données moyennées entre 2015 et 2020, le Canada se classe cinquième en ce qui concerne la pénétration nationale des compétences en IA, au-dessus de la moyenne du G20 (neuvième), mais derrière l'Inde, les États-Unis, la Chine et l'Allemagne.

Il faudra décider s'il est nécessaire d'élargir les programmes de diplôme universitaire et à quel niveau d'études. Bien que certains pensent que les doctorats devront être plus interdisciplinaires (p. ex. exiger des biologistes qu'ils comprennent non seulement la biologie, mais également l'apprentissage machine et d'autres concepts informatiques), d'autres croient que les programmes de premier cycle seront élargis alors que les programmes de doctorat resteront davantage axés sur les compétences disciplinaires approfondies (Fleming, 2018). Les universités ne

sont pas les seules responsables de la formation des futurs développeurs en IA. Les leaders du domaine tirent aussi parti de l'industrie pour soutenir les initiatives de formation pratique d'une façon qui bénéficie aux étudiants, aux entreprises et, en fin de compte, à l'écosystème d'IA canadien (encadré 4.1).

La formation des futurs chercheurs de l'ère de l'IA ne se limitera pas à élargir leurs connaissances et leurs compétences techniques. Les universitaires admettent la nécessité d'enseigner aux futurs scientifiques comment penser les dilemmes éthiques, culturels et sociaux découlant du développement de l'utilisation de l'IA

(Dignum, 2020a). Par exemple, l'Université de Toronto lancera un programme pilote qui intégrera des modules éthiques dans son programme de premier cycle en informatique, et réorganisera ses cours de manière à apprendre aux étudiants à déceler les questions éthiques et les conséquences sociales générales des technologies de demain (UofT, 2021b). Conformément à la stratégie du programme Embedded EthiCS de l'Université Harvard, l'objectif n'est pas nécessairement d'améliorer la pensée abstraite chez les scientifiques et les ingénieurs en informatique, mais plutôt d'enseigner aux étudiants à inclure les facteurs éthiques dans la conception de nouvelles technologies. Comme Bezuidenhout et Ratti (2020) le suggèrent, la modification des programmes d'études visant à mettre l'accent sur les questions auxquelles feront face les scientifiques des données dans leurs activités professionnelles peut influencer le développement du caractère par l'internalisation d'une formation à l'éthique susceptible d'être appliquée au contexte microéthique.

Encadré 4.1 Mitacs et Scale AI

Mitacs est un organisme sans but lucratif canadien qui offre des stages financés et des possibilités de bourse, et met en relation les étudiants, les entreprises et le secteur public, créant ainsi des occasions de formation postsecondaire permettant aux récents diplômés de développer leur expertise pratique. L'organisme s'est engagé à soutenir l'écosystème de l'IA en mettant en relation les petites et moyennes entreprises (PME) qui souhaitent lancer des projets d'IA avec des étudiants de deuxième cycle et postdoctoraux capables d'offrir de l'expertise et des conseils sur la façon de commencer.

Scale AI, une supergrappe d'IA canadienne, investit également dans la formation sur l'IA au Canada. En subventionnant des programmes de formation agréés destinés aux travailleurs à temps plein, elle cherche à améliorer les compétences actuelles de la main-d'œuvre en activités numériques. Ces programmes consacrés à susciter l'intérêt en sciences, en technologie, en génie et en mathématiques (STGM) chez les jeunes procurent aussi des investissements dans la main-d'œuvre de demain.

La structure organisationnelle actuelle du financement de la recherche, qui divise celle-ci en santé, en sciences et en génie, et en sciences sociales et sciences humaines, conformément au mandat des trois organismes, devra être adaptée pour « encourager, faciliter, évaluer et soutenir la recherche multidisciplinaire » [traduction libre] (Naylor *et al.*, 2017). Un exemple d'un tel ajustement est le lancement du fonds Nouvelles frontières en recherche des trois organismes, qui vise à appuyer la recherche interdisciplinaire au Canada (CRSH, 2021).

Les conséquences sociales de la recherche employant l'IA

- 5.1 Confiance dans l'IA : biais d'automatisation et aversion aux algorithmes
- 5.2 Gérer le biais et la discrimination dans les systèmes d'IA
- 5.3 Impact de l'IA sur les communautés autochtones
- 5.4 L'EDI dans la recherche sur l'IA et dans l'accès à la technologie d'IA
- 5.5 Répercussion sur le marché du travail
- 5.6 Répercussions environnementales des systèmes d'IA
- 5.7 Sécurité de l'IA

Constatations du chapitre

- Un manque de confiance dans l'IA utilisée dans d'autres domaines pourrait gêner son adhésion en sciences et en génie. Inversement, une confiance dénuée de toute critique dans l'IA pourrait mener à une dépendance excessive aux résultats de la recherche produits par cette technologie.
- Étant donné le problème bien connu de la perpétuation par les systèmes d'IA du biais et de la discrimination contre des groupes historiquement marginalisés, il sera essentiel que son utilisation dans la recherche en sciences et en génie ne produise pas des résultats biaisés susceptibles d'aboutir à des conséquences discriminatoires.
- Les importantes inégalités actuelles dans l'accès aux ressources, à l'infrastructure et aux possibilités d'emploi relatives à l'IA pourraient avoir une incidence négative sur les progrès en EDI dans la recherche en sciences et en génie employant l'IA.
- Bien que la nature et l'ampleur des répercussions de l'IA sur la main-d'œuvre en sciences et en génie demeurent floues, de nombreux emplois seront transformés par sa présence accrue.
- Les systèmes d'IA peuvent avoir un gros impact environnemental en raison de l'énergie requise pour alimenter l'infrastructure de calcul. Cependant, l'atténuation de ces impacts fait l'objet d'une étude constante et l'IA est prometteuse en ce qui concerne la recherche en sciences et en génie susceptible d'aider à réduire les dommages environnementaux et les changements climatiques.
- Bien que l'on admette que de solides mesures de sécurité sont nécessaires pour protéger le système d'IA contre diverses sortes d'attaques, peu de recherches portent spécifiquement sur le rôle de la cybersécurité dans le contexte de l'IA en sciences et en génie.

Il ne fait aucun doute que l'intensification de l'utilisation de l'IA dans la recherche en sciences et en génie influera sur la société, l'économie et l'environnement. La confiance sociale dans la technologie d'IA pourrait nuire à son adhésion en sciences et en génie, et son utilisation responsable exigera de gérer les risques de perpétuer le biais ou la discrimination en ce qui concerne les constatations produites par les systèmes d'IA et l'accès à la technologie elle-même. Le recours accru à l'IA en sciences et en génie pourrait également influencer sur les marchés du travail et sur la demande en compétences dans ces domaines, et accroître les émissions de gaz à effet de serre (GES). De plus, des mesures de cybersécurité seront nécessaires afin de garantir l'utilisation sécuritaire responsable de l'IA dans ces domaines.

Beaucoup ou la plupart de ces questions — y compris la confiance sociale, le biais et la discrimination, l'accès équitable aux ressources, la perturbation du marché du travail, la dégradation de l'environnement et la sécurité — sont des défis sociaux qui datent d'avant l'introduction générale de l'IA dans divers secteurs de la société et son utilisation dans le contexte particulier de la recherche en sciences et en génie. L'arrivée de l'IA devrait donc être vue comme un phénomène qui amplifie les problèmes préexistants, tout en les complexifiant et en créant de nouveaux défis connexes.

5.1 Confiance dans l'IA : biais d'automatisation et aversion aux algorithmes

Un manque de confiance dans l'IA pourrait entraver son adoption en sciences et en génie. Par exemple, O'Connor *et al.* (2019) suggère que le manque de confiance dans l'IA est actuellement un obstacle à son emploi dans les examens systématiques de la recherche. Pour le surmonter, les systèmes d'IA devront bâtir des données probantes fiables en démontrant de façon transparente des résultats fructueux et reproductibles. La confiance est également de mise pour une collaboration humain-machine efficace, sans laquelle les membres de l'équipe pourraient avoir à « consacrer une énergie et un temps inutiles à réinspecter les travaux et à revalider les décisions » [traduction libre] (Hou *et al.*, 2021).

Les découvertes scientifiques ou les concepts de génie produits par l'IA peuvent être injustement perçus comme supérieurs ou inférieurs à leurs équivalents d'origine humaine

Les différences d'exactitude ou d'utilité perçues de l'IA en comparaison du travail humain en sciences et en génie peuvent démontrer un *biais d'automatisation* (Lopez *et al.*, 2019), selon lequel les résultats produits par la machine sont vus comme plus dignes de confiance que ceux produits par les humains, ou une *aversion aux algorithmes* (Dietvorst *et al.*, 2015), où les résultats générés par l'IA sont perçus comme moins dignes de confiance que ceux produits par les humains.

Une étude réalisée par Lopez *et al.* (2019), par exemple, a constaté que les gens étaient quelque peu plus enclins à voir les croquis de bateau conçus par l'IA comme plus fonctionnels que ceux dessinés par l'humain. Cet effet était exacerbé quand les croquis étaient explicitement étiquetés comme ayant été produits par un ordinateur ou par un humain, ce qui porte à croire que la fonctionnalité perçue peut démontrer un biais d'automatisation. Similairement, une recherche effectuée par Logg *et al.* (2018) a révélé que les individus attribuaient constamment une plus grande valeur à un même conseil quand il était étiqueté comme émanant d'un algorithme, que quand il était étiqueté comme formulé par un humain. En revanche, une recherche réalisée par Dietvorst *et al.* (2015) a démontré que les gens sont bien plus enclins à perdre

confiance dans des résultats d'origine algorithmique que dans les résultats d'origine humaine après avoir vu les erreurs système, même quand ils ont constaté que les algorithmes étaient généralement plus performants que les humains. En fait, la confiance des participants dans le modèle algorithmique baissait invariablement une fois qu'ils avaient vu qu'il faisait de petites erreurs, alors que la constatation de grosses erreurs faites par les humains ne réduisait pas invariablement leur confiance dans l'humain (Dietvorst *et al.*, 2015).

Fait notable, tant dans l'étude de Lopez *et al.* (2019) que dans celle de Dietvorst *et al.* (2015), l'IA était invariablement plus performante que ses concurrents humains, ce qui signifie que dans les deux cas, la confiance dans cette technologie était la meilleure stratégie pour garantir l'exactitude. Cependant, l'exactitude seule peut être insuffisante pour bâtir la confiance dans l'IA; la transparence et l'explicabilité des résultats (section 3.1.2) peut être essentielle (Hou *et al.*, 2021). De plus, Logg *et al.* (2018) ont constaté que les gens étaient plus prêts à suivre un conseil formulé par un algorithme qu'un conseil humain, même dans les cas où le système était une boîte noire, et suggèrent que, pour les profanes, davantage de renseignements pourraient davantage accroître les doutes qu'augmenter la confiance. Néanmoins, les cadres de confiance dans l'IA élaborés par la Commission européenne (2019) et l'OCDE (2021a) (tableau 5.1) considèrent la transparence et l'explicabilité comme des éléments vitaux de la confiance dans l'IA, en particulier dans la prise de décision automatisée. L'interprétabilité sera probablement nécessaire pour que les gens fassent confiance à l'IA dans les contextes décisionnels touchant les sciences et le génie, tels que les décisions de financement et les examens par les pairs (section 4.1).

L'IA peut être perçue comme digne de confiance ou non pour différentes raisons dans différents contextes

La confiance en l'IA se heurte à différents écueils selon le contexte (AI HLEG, 2019). Par exemple, les systèmes d'IA qui formulent des recommandations musicales ne créent pas les mêmes inquiétudes à propos de la confiance que ceux qui recommandent des traitements ou des soins médicaux critiques dans la prise de décision gouvernementale. Quoi qu'il en soit, même des systèmes d'IA apparemment inoffensifs peuvent soulever des craintes concernant la confiance touchant leurs pratiques de recueil et d'utilisation des données, en particulier quand ces données sont utilisées par d'autres systèmes d'IA. Par conséquent, la confiance dans l'IA doit être abordée selon le contexte unique de son utilisation dans la recherche et la découverte en sciences et en génie. Cependant, cette question ne fait que peu ou pas l'objet de recherches pour ces domaines particuliers. Néanmoins, la méfiance de la population dans l'IA dans d'autres domaines — comme le système de justice, les soins de santé, le marché du travail

et les réseaux sociaux — pourrait nuire à la perception que l’IA en sciences et en génie est digne de confiance. Si c’est le cas, bâtir la confiance dans l’IA dans ces deux disciplines peut nécessiter de bâtir la confiance dans l’IA dans ces autres domaines publics.

Tableau 5.1 Principes pour une IA digne de confiance

Commission européenne	OCDE
<ul style="list-style-type: none"> • Reddition de comptes • Diversité, non-discrimination et équité • Bien-être environnemental et sociétal • Facteur et supervision humains • Respect de la vie privée et gouvernance des données • Solidité et sécurité techniques • Transparence 	<ul style="list-style-type: none"> • Reddition de comptes • Valeurs centrées sur l’humain et équité • Croissance inclusive, développement durable et bien-être • Solidité et sécurité • Transparence et explicabilité

(AI HLEG, 2019; OCDE, 2021a)

L’OCDE et la Commission européenne ont toutes deux élaboré des cadres pour une IA digne de confiance. Les deux entités définissent des méthodes et des outils techniques et non techniques pouvant être utilisés pour mettre en application ces principes.

5.2 Gérer le biais et la discrimination dans les systèmes d’IA

L’utilisation responsable de l’IA en sciences et en génie doit éviter de perpétuer le biais et la discrimination contre des individus ou des groupes. Comme il est noté au chapitre 3, les résultats discriminatoires sont un problème bien connu de l’IA, et son emploi dans divers contextes s’est révélé discriminatoire à l’encontre de groupes historiquement marginalisés. Par exemple, on a découvert que les systèmes d’IA étaient discriminatoires à l’égard des femmes (Dastin, 2018) et des personnes souffrant de déficience mentale (Fruchterman et Mellea, 2018) dans les décisions d’embauche. L’utilisation de l’IA a également créé de la discrimination à l’encontre des Noirs en soins de santé (Obermeyer *et al.*, 2019) et dans les évaluations du risque de récidive dans le système de justice pénale (Dressel et Farid, 2018). On a découvert que l’IA était discriminatoire à l’endroit des étudiants provenant d’écoles historiquement peu performantes dans l’attribution des grades (Satariano, 2020), et des Noirs (Sap *et al.*, 2019) et des personnes souffrant de handicap (Hutchinson *et al.*, 2020) dans les tentatives de détection des discours jugés toxiques ou haineux. Ces types de biais discriminatoires pourraient

constituer de nouveaux défis pour les lois antidiscrimination au Canada, par exemple en matière d'accès aux données confidentielles dans les cas de litige concernant le biais et la discrimination causés par l'IA. Le chapitre 6 examine les questions légales connexes.

Le risque que l'IA soit discriminatoire en sciences et en génie dépend du but de son utilisation

Généralement, l'emploi de l'IA en sciences et en génie ne touche pas directement les individus ou les groupes, du moins en ce qui a trait à l'objet de la *découverte* et pas de la *prise de décision*. Cependant, comme il est noté au chapitre 4, l'IA peut en fait servir pour la prise de décision dans les domaines des sciences et du génie, par exemple pour l'examen par les pairs et pour les décisions de financement. Dans le contexte de la découverte, il sera important de veiller à ce que la conception et l'utilisation des systèmes d'IA ne nuisent pas aux individus indirectement, par exemple en produisant des résultats inexacts qui seront utilisés par certains pour justifier des décisions discriminatoires. Bien entendu, ce problème se pose également pour les résultats en sciences et en génie non issus de l'IA. Cependant, l'opacité de certains systèmes d'IA signifie que le processus qui a produit les résultats en question peut ne pas être transparent pour les personnes qui utilisent les constatations pour prendre des décisions, pas même pour les chercheurs qui ont conçu ces systèmes (section 3.1.2). Par conséquent, on doit faire en sorte que la diffusion responsable et éthique des résultats de constatations de recherche produites par la machine (section 3.2.2) ne perpétue pas la discrimination.

Les biais dans les données d'apprentissage peuvent perpétuer la discrimination, mais il est possible d'améliorer ces données en accordant une plus grande importance à leur provenance, à la reddition des comptes et à la transparence

Les résultats discriminatoires traduisent souvent des biais dans les données d'apprentissage du modèle d'IA, qui peuvent être le reflet de biais historiques. Dans d'autres cas, la discrimination peut découler de l'exclusion de classes d'individus qui ne produisent pas de grandes quantités de données, comme les personnes vivant en zone rurale dans les pays à faibles revenus. De plus, les données pourraient être simplement de mauvaise qualité ou être criblées d'erreurs (WEF, 2018). Toutefois, les biais historiques et d'échantillonnage ne sont pas les seuls biais touchant les ensembles de données; par exemple, les données subissent souvent divers traitements avant d'être utilisées dans les applications d'IA, et les décisions subjectives que comprend un tel traitement peuvent avoir des conséquences importantes sur le comportement des systèmes d'IA (Veale et Binns, 2017).

Un moyen suggéré par les experts de gérer les biais et d'éviter les résultats discriminatoires dans la recherche fondée sur l'IA et de veiller à la *provenance des données*, c'est-à-dire documenter l'historique et le processus de sélection d'un ensemble de données, de sa construction et de son utilisation (WEF, 2018; West et al., 2019). Pour opérationnaliser ce processus, Gebru et al. (2021) ont proposé des *fiches d'ensembles de données*, qui documenteraient les diverses caractéristiques des ensembles de données, telles que le motif de leur création, leur composition,



« Ces choix ne sont pas neutres, mais plutôt inévitablement chargés de valeurs et reflètent la philosophie des individus qui conçoivent le modèle. »

leur processus de recueil, leur prétraitement et leur étiquetage ou leur utilisation recommandée (y compris la distribution et la tenue à jour).

Cela aiderait les créateurs et les utilisateurs des ensembles de données en encourageant les créateurs à s'engager dans « une réflexion approfondie sur les processus de création, de diffusion et de tenue à jour d'un ensemble de données, y compris sur les hypothèses sous-jacentes, risques de préjudice et conséquences de l'utilisation » [traduction libre] (Gebru et al., 2021), tout en fournissant également aux utilisateurs les renseignements nécessaires pour

prendre une décision éclairée quant à l'utilisation de ces données. Les fiches d'ensembles de données peuvent également être utiles aux « décideurs, aux organismes de défense des consommateurs, aux individus dont les données sont incluses dans ces ensembles et aux personnes susceptibles d'être touchées par les modèles entraînés ou évalués à partir de ces ensembles de données » [traduction libre] (Gebru et al., 2021). La façon dont les fiches d'ensembles de données pourraient faciliter la reproductibilité de la recherche en IA est examinée à la section 3.1.

Cependant, la provenance des données seules ne résoudra pas les problèmes de discrimination en IA. Même dans les cas où les modèles d'IA sont entraînés à partir d'ensembles de données de grande qualité et non biaisés, ils peuvent encore produire des résultats discriminatoires s'ils ne sont pas appropriés à l'étude, si leurs paramètres sont involontairement discriminatoires ou si la supervision humaine ou la transparence est insuffisante, voire en cas de discrimination intentionnelle (WEF, 2018). Par conséquent, en plus de la provenance des données, la documentation des choix subjectifs effectués par les chercheurs durant le développement d'un modèle peut aider à déceler les biais potentiellement discriminatoires (Selbst et Barocas, 2018). En fait, ces choix ne sont pas neutres, mais plutôt inévitablement chargés de valeurs et reflètent la philosophie des individus qui conçoivent le modèle (Veale et Binns, 2017). Par conséquent, la reddition de comptes est cruciale pour déceler et atténuer les résultats discriminatoires (Berendt et Preibusch, 2017).

Les chercheurs qui travaillent dans des domaines tels que l'exploration de données sensibles à la discrimination et l'équité, la reddition de comptes et la transparence dans l'apprentissage machine ont proposé diverses techniques de



« Le biais et la discrimination dans les systèmes d'IA sont souvent le résultat de problèmes systémiques de pouvoirs et d'inégalité dans les établissements qui produisent ces systèmes, un défi exacerbé par le manque global. »

calcul et démarches techniques pour déceler et atténuer les biais discriminatoires dans les ensembles de données et dans les systèmes d'IA (Veale et Binns, 2017)¹². Cependant, les mesures techniques seules ne sont pas suffisantes pour éviter la discrimination; les actions et les choix humains dans le développement et l'utilisation de l'IA peuvent avoir des conséquences discriminatoires, même après que les ensembles de données ont fait l'objet d'une recherche de biais (Berendt et Preibusch, 2017).

Enfin, l'IA peut aussi produire des résultats discriminatoires en raison du manque de diversité dans le développement du système; comme West *et al.* (2019) l'expliquent, « s'attaquer aux défis des biais dans les systèmes techniques nécessite de s'attaquer à la diversité de la main-d'œuvre, et

inversement » [traduction libre]. Le biais et la discrimination dans les systèmes d'IA sont souvent le résultat de problèmes systémiques de pouvoirs et d'inégalité dans les établissements qui produisent ces systèmes (West *et al.*, 2019), un défi exacerbé par le manque global de diversité de genre et de race dans le domaine de la recherche en IA (section 5.4).

5.3 Impact de l'IA sur les communautés autochtones

Le mouvement de souveraineté des données autochtones vise à accroître l'autodétermination et la gouvernance des peuples autochtones en ce qui concerne les données et à corriger les biais dans les données susceptibles de causer de la discrimination à leur rencontre

Au Canada et ailleurs dans le monde, les communautés autochtones sont touchées de façon disproportionnée par les biais dans les ensembles de données. Par exemple, le manque de données désagrégées sur la santé au Canada obscurcit les différences dans les résultats en matière de santé pour les peuples autochtones (BCOHR, 2020),

12 Stoyanovich *et al.* (2016) et Hajian *et al.* (2016) ont examiné en détail certaines de ces techniques.

alors que les données socioéconomiques sur les populations autochtones vivant en zone urbaine ou dans des petites communautés sont souvent faussées en raison



« Au Canada et ailleurs dans le monde, les communautés autochtones sont touchées de façon disproportionnée par les biais dans les ensembles de données. »

de la taille relativement faible des échantillons (Steffler, 2016). Cette situation a engendré des mouvements lancés par les Autochtones pour la souveraineté des données les concernant, soit le droit à l'autodétermination et à la gouvernance relatives à l'accès, à la propriété et à l'utilisation des données et des analyses de données (Rainie *et al.*, 2017; Crawford *et al.*, 2019). Le concept de souveraineté des données autochtones suscite de plus en plus l'intérêt à mesure que « les mégadonnées, les données ouvertes, la science ouverte et la réutilisation des données deviennent une partie intégrante de la recherche et des

pratiques institutionnelles » [traduction libre] (Carroll *et al.*, 2020). Parmi les organismes et les initiatives promouvant cette souveraineté, citons :

- le U.S. Indigenous Data Sovereignty Network, qui a été mis sur pied en 2016 dans le but de relier les utilisateurs de données, les chefs tribaux, les fournisseurs de technologies de l'information et des communications, les chercheurs, les décideurs et les planificateurs, les entreprises, les fournisseurs de services et les défenseurs de la communauté indiens américains, autochtones de l'Alaska et autochtones d'Hawaii, afin de mettre en commun leurs initiatives de données, leurs succès et défis et leurs ressources » [traduction libre] (USIDSN, s.d.);
- l'initiative Local Contexts, qui a été lancée en 2010 afin de faciliter la gestion des questions concernant les données autochtones dans les environnements numériques, et qui se concentre sur « l'accroissement de la participation des Autochtones dans la gouvernance des données par l'intégration des valeurs autochtones dans les systèmes de données » [traduction libre] (Local Contexts, s.d.);
- la Global Indigenous Data Alliance (GIDA), fondée en 2020, qui est « un réseau de chercheurs, de spécialistes des données et d'activistes politiques autochtones qui effectuent la promotion de la souveraineté des données autochtones dans leur État-nation et au niveau international » [traduction libre] (GIDA, s.d.s.d.-a). La GIDA travaille également à la promotion des principes CARE pour la gouvernance des données autochtones (GIDA, s.d.-b);

- le Centre de gouvernance de la formation des Premières Nations (CGPIN), établi au Canada en 2009, et spécialement mandaté par les Chefs en assemblée de l'Assemblée des Premières Nations (Résolution no 48, décembre 2009) » (CGPIN, s.d.). Le CGPIN appuie la souveraineté des données autochtones et « le développement de la gouvernance et de la gestion de l'information à l'échelle communautaire grâce à des partenariats régionaux et nationaux » (CGPIN, s.d.). Le CGPIN s'emploie principalement à concevoir des initiatives de recueil des données dans les communautés des Premières Nations. Il a également élaboré une norme sur la propriété, le contrôle, l'accès et la possession (PCAP) des données sur la façon de conduire de la recherche avec les Premières Nations;
- la Stratégie nationale inuite sur la recherche, élaborée par l'Inuit Tapiriit Kanatami en 2018, qui promeut l'amélioration de la gouvernance inuite dans la recherche, et notamment l'accès aux données sur les Inuits, leur propriété et leur contrôle (ITK, ss.d.).

Les perspectives autochtones sur l'IA peuvent constituer un apport précieux qui remettra en question les hypothèses et biais culturellement dominants

En plus de travailler à la souveraineté des données autochtones, les chercheurs autochtones — comme les membres de l'Indigenous Protocol and Artificial Intelligence Working Group — étudient la relation entre l'IA et les perspectives autochtones. En 2020, ce groupe de travail a publié un exposé de position examinant une variété de points de vue, de systèmes de connaissances et de pratiques technologiques autochtones concernant l'IA (Lewis, 2020). Ce document étudie notamment le contrôle et la souveraineté autochtone sur le matériel, les logiciels et les données; le rôle du savoir traditionnel dans les systèmes d'IA et la nécessité de protéger ce savoir; la conception de systèmes d'IA conformément au cadre éthique autochtone; les dangers que pose l'IA pour les communautés autochtones en tant qu'extension des pratiques coloniales d'exploitation, d'extraction et de contrôle; le rôle des langues autochtones en IA et la relation entre la langue et les processus de calcul; et la nature de la relation entre les êtres humains et l'IA (Lewis, 2020).

Ce groupe de travail se focalise sur la façon dont les épistémologies et les ontologies autochtones peuvent guider le développement de l'IA. Il cherche à offrir une vision différente des « épistémologies rationalistes occidentales » à la base du développement actuel de l'IA et des préjugés et biais qu'elles peuvent receler (Lewis, 2020). Les points de vue autochtones sur l'IA peuvent servir à « remettre en question l'anthropocentrisme fondamental de la science et technologie occidentale [en offrant] des épistémologies qui refusent de centrer ou d'élever l'humain [en faveur d'un développement technologique basé sur] les principes et les pratiques la durabilité sociale et environnementale » [traduction libre] (Lewis, 2020).

Les principes CARE pour la gouvernance des données autochtones peuvent faciliter l'utilisation éthique de ces données

Les principes CARE (bénéfice collectif, pouvoir de contrôle, responsabilité et éthique) relatifs à la gouvernance des données autochtones, qui sont un exemple de principes de gestion et de gestion des données facilitant le partage et l'utilisation responsables et éthiques des données (section 3.1.4), ont été élaborés par l'Indigenous Data Sovereignty Group, au sein du RDA. Ces principes complètent les actuels principes FAIR (données faciles à trouver, accessibles, interopérables et réutilisables) sur la gestion et la gouvernance des données scientifiques et sont conçus pour être mis en œuvre conjointement (Carroll *et al.*, 2020). Les principes CARE sont une réponse directe aux principes FAIR de gouvernance des données — qui peuvent ignorer le déséquilibre des pouvoirs et le contexte historique — et visent à reconnaître le rôle des données dans la promotion de l'innovation et de l'autodétermination autochtones (Carroll *et al.*, 2020; GIDA, s.d.-b). Fait important, ces principes ne s'appliquent pas seulement au savoir autochtone ou traditionnel, mais également aux données scientifiques (Carroll *et al.*, 2021).

5.4 L'EDI dans la recherche sur l'IA et dans l'accès à la technologie d'IA

L'accès à la technologie d'IA soulève la question de l'EDI. Bien qu'on dispose de quelques données sur la diversité de genre des chercheurs dans le domaine de l'IA, on manque globalement de renseignements sur d'autres facteurs démographiques, comme l'âge, la race ou l'ethnicité, le handicap, l'orientation sexuelle et le statut socioéconomique (Young *et al.*, 2021). L'interrelation entre l'EDI et l'IA dans le système de financement de la recherche est examinée au chapitre 4.

Le domaine de la recherche en IA manque globalement de diversité de genre et de race

En 2018, moins de 22 % des professionnels des domaines de l'IA et de la science des données dans le monde étaient des femmes, approximativement 24 % au Canada (WEF, 2018)¹³. En intelligence artificielle, les femmes ont plus de probabilité de détenir un emploi moins prestigieux et rémunéré que les hommes (p. ex. généralement en analyse, en préparation et en exploration des données, plutôt qu'à des postes plus prestigieux en génie et en apprentissage machine), sont sous-représentées dans des secteurs d'activité exigeant davantage de compétences techniques et présentent des taux de roulement et d'attrition supérieurs à ceux des hommes (Young *et al.*, 2021). Les femmes représentent environ 15 % des auteurs

¹³ Remarque : Cette étude et plusieurs de celles mentionnées dans cette section ne font pas la distinction entre le genre et le sexe et peuvent utiliser des termes tels que « woman » (genre) et « female » (sexe) de façon interchangeable. Le rapport utilise alors la terminologie correspondant au genre.

d'articles sur l'IA publiés dans arXiv (Element AI, 2020) et seulement environ 18 % des auteurs dans les plus grandes conférences sur l'IA (Element AI, 2019). De même, seulement 16 % des postes de la filière de permanence dont l'axe de recherche principale est l'IA sont occupés par des femmes (HAI, 2021)¹⁴ et seulement 15 % du personnel de recherche en IA chez Facebook et 10 % chez Google sont des femmes (Simonite, 2018). Les femmes sont également sous-représentées dans le brevetage : à l'échelle mondiale, on relève une femme pour trois hommes dans le brevetage en IA, une proportion encore plus faible au Canada, avec une femme pour six hommes (ISDE, 2019a). En revanche, il n'existe pas de données sur les travailleurs transgenres et autres minorités sexuelles (West *et al.*, 2019).

Bien qu'il existe peu de données au sujet de la race ou de l'ethnicité en IA, les renseignements disponibles indiquent que ce domaine affiche généralement une faible diversité raciale. Par exemple, la répartition raciale et ethnique des nouveaux titulaires d'un doctorat en IA résidant aux États-Unis est d'environ 46 % de blancs, 22 % d'Asiatiques, 3 % d'Hispaniques, de 2 % de Noirs et 2 % de personnes multiraciales (la race et l'ethnicité étant inconnues dans 25 % des cas) (HAI, 2021). Bien qu'on ne dispose pas de données sur la diversité raciale ou ethnique des chercheurs en IA dans le secteur privé, seul un faible pourcentage de la main-d'œuvre totale dans les grandes compagnies technologiques est d'origine noire ou hispanique — 2,5 et 3,6 % chez Google, 4 et 5 % chez Facebook et 4 et 6 % chez Microsoft, respectivement (West *et al.*, 2019). Des efforts sont actuellement déployés pour accroître cette diversité, par exemple, par Black in AI, une organisation qui œuvre à accroître la représentation des Noirs dans le domaine. Celle-ci a tenu plusieurs ateliers dans les plus grandes conférences sur l'IA et a aidé à accroître notablement le nombre de participants noirs à ces conférences (HAI, 2021).

Il existe d'importantes inégalités dans l'accès à l'IA et aux mégadonnées pour la recherche

Il y a actuellement une forte inégalité dans la répartition des ressources, de l'infrastructure et des compétences en matière de production, de diffusion et l'utilisation des mégadonnées pour la recherche scientifique (Leonelli, 2020). On appelle généralement fossé numérique l'écart de possibilités entre les personnes qui ont accès aux technologies de l'information et des communications et à Internet et celles qui n'y ont pas accès (Carter *et al.*, 2020). Dans le contexte de l'IA, ce fossé — que Yu (2020) appelle le *fossé algorithmique* — renvoie aux différences entre les individus, les chercheurs, les établissements, les organismes et les

14. Une étude réalisée par le Forum économique mondial en arrive à une conclusion quelque peu différente sur la diversité de genre en IA aux études supérieures, et constate que l'IA est un des seuls secteurs dans le monde dans lequel les femmes sont plus nombreuses que les hommes dans le bassin de talents (WEF, 2018). Cependant, West *et al.* (2019) critiquent la méthodologie de cette étude, c'est-à-dire le fait qu'elle utilise LinkedIn comme source de données primaire et qu'elle attribue le genre à partir du prénom.

régions géographiques en ce qui concerne l'accès à l'IA, son utilisation et ses résultats. Le fossé algorithmique empêche une partie importante de la population de profiter des avantages de la technologie d'IA, et il continue à s'élargir (Leonelli, 2020; Yu, 2020). De plus, il nuit aux constatations produites par la recherche guidée par les données. Par exemple, il limite la disponibilité des données sur certains groupes et certaines régions géographiques, et contribue à ce que des données soient largement diffusées pendant que d'autres demeurent confidentielles à cause de facteurs tels que la valeur commerciale (Leonelli, 2020).



« Au Canada, les investissements publics dans le secteur de l'IA profitent d'ailleurs principalement au secteur privé. »

Ce fossé est déterminé et façonné par des facteurs sociaux comme la démographie, la culture, les politiques et la réglementation; par des facteurs techniques comme l'infrastructure, les algorithmes et les données d'apprentissage; et par des facteurs sociotechniques tels que les compétences et les connaissances informatiques, ainsi que les perceptions et les croyances concernant l'IA (Carter *et al.*, 2020). Les facteurs économiques peuvent

également exacerber le fossé algorithmique pour la recherche scientifique. Par exemple, en raison du coût élevé des ressources de calcul et de la compétition pour attirer les talents en recherche en IA, les capacités sont souvent concentrées dans les entreprises privées plutôt que dans les universités ou dans le secteur public (OCDE, 2018b). Comme l'affirme l'OCDE (2018a), « cela peut conduire à la concentration de la découverte scientifique et soulever des craintes d'un monopole excessif du savoir scientifique » [traduction libre]. Au Canada, les investissements publics dans le secteur de l'IA profitent d'ailleurs principalement au secteur privé (Brandusescu *et al.*, 2021).

Au Canada, le fossé numérique est plus apparent dans les régions rurales et nordiques du pays, qui souffrent souvent d'un gros déficit de connectivité à Internet (CAC, 2021b). Par exemple, moins de la moitié (48 %) des personnes vivant à l'extérieur des grandes agglomérations canadiennes (c.-à-d. les villes dont la population est supérieure à 10 000 habitants) ont accès à Internet haute vitesse (c.-à-d. débit supérieur à 50 Mbits/s), comparé aux quarts (76 %) des gens qui vivent dans ces agglomérations (StatCan, 2021b). De même, par comparaison avec la moyenne canadienne (92 %), on retrouve de plus faibles taux d'utilisation d'Internet chez les peuples autochtones (88 %), chez les personnes souffrant de handicap (85 %), chez les personnes sans emploi (85 %) et chez les personnes de plus de 75 ans (62 %) (StatCan, 2021a). Il n'existe que peu ou pas de renseignements disponibles sur le fossé algorithmique au Canada, par exemple sur la proportion de chercheurs du secteur public ou du secteur privé ayant accès à la technologie d'IA ou sur leur profil démographique. Ces renseignements seraient utiles pour élaborer des politiques facilitant l'utilisation de l'IA en sciences et en génie au Canada.

Les iniquités actuelles dans le système de recherche en IA peuvent nuire aux résultats en recherche en général

Comme il est noté à la section 5.2, il est important d'accroître la diversité dans la recherche sur l'IA pour déceler et atténuer les biais discriminatoires potentiels en intelligence artificielle. Cependant, la valeur de la diversité en IA dépasse les préoccupations éthiques et sociales; concernant l'équité, et la diversité dans la recherche utilisant l'IA peut améliorer les résultats épistémiques en sciences et en génie. Dans son rapport publié en 2012, le comité d'experts sur les femmes dans la recherche universitaire du CAC a constaté qu'« élargir le contingent de chercheurs au Canada en ouvrant des possibilités aux femmes et aux autres groupes sous-représentés pourrait permettre d'obtenir des résultats de recherche plus robustes » (CAC, 2012). Des études ultérieures ont constaté que les chercheurs appartenant aux groupes sous-représentés dans leur discipline ont tendance à produire plus de nouveautés scientifiques et à établir plus de nouveaux liens conceptuels (Hofstra *et al.*, 2020) et que la diversité de genre peut conduire à améliorer la science (Nielsen *et al.*, 2017).

D'autres études ont relevé un lien entre la diversité cognitive — c'est-à-dire la diversité de points de vue ou de style traitement de l'information — et l'amélioration de la résolution de problème dans les équipes (Reynolds et Lewis, 2017). La diversité cognitive peut également être bénéfique sur le plan épistémique pour la recherche scientifique en stimulant la créativité scientifique et en conduisant les chercheurs à rechercher et à employer des démarches, des méthodes de recherche, des types de données probantes, des hypothèses et des théories nouveaux (Rolin, 2019). De plus, de nombreux philosophes des sciences et épistémologues sociaux avancent que, dans les bonnes conditions, l'augmentation de la diversité démographique peut simultanément accroître la diversité cognitive et donc apporter des bénéfices épistémiques à la recherche scientifique (Fehr, 2011; Rolin, 2019).

5.5 Répercussion sur le marché du travail

Bien que les répercussions de l'IA sur le marché du travail aient été largement étudiées (voir OCDE (2021b)), il existe comparativement peu de renseignements sur son incidence sur l'emploi en sciences et en génie. Cependant, une analyse sur le sujet effectuée par Webb (2020) a constaté que les métiers les plus exposés à la suppression sous l'effet de l'IA étaient ceux d'ingénieur en métallurgie et en matériaux, d'ingénieur chimiste, de physicien et d'astronome, de scientifique de l'atmosphère et de l'espace, de technicien en génie et en sciences, ainsi que les autres professionnels des sciences et du génie. Il est probable que l'IA aura un impact différent sur le marché du travail dans différents secteurs des sciences du génie. Par exemple, les champs de recherche qui comportent davantage d'activités manuelles, comme la science animale et l'archéologie, pourraient être moins

exposés à cette technologie (Brynjolfsson *et al.*, 2018), tout comme ceux qui incluent le raisonnement sur des situations nouvelles (Webb, 2020).

L'ampleur de l'influence qu'aura l'IA sur le marché du travail en sciences et en génie au Canada n'est pas claire. En 2016, la taille du marché du travail dans la catégorie « Sciences naturelles et appliquées et domaines apparentés » de la Classification nationale des professions était de 1,27 million de personnes, soit près de 7 % de la main-d'œuvre totale. Environ 725 000 de ses individus appartenaient à la sous-catégorie « Personnel professionnel des sciences naturelles et appliquées » et 549 000 à « Personnel technique assimilé aux sciences naturelles et appliquées » (StatCan, 2016). L'IA pourrait également avoir une incidence sur les emplois en aval de la recherche en sciences et en génie, comme ceux concernés par la commercialisation des découvertes. Cependant, il existe très peu de données probantes à propos de l'impact de l'IA sur ces métiers. L'utilisation de l'IA en sciences et en génie pourrait indirectement créer de nouveaux emplois dans ces secteurs en raison de son potentiel à produire des innovations et des percées scientifiques (Lane et Saint-Martin, 2021).

L'IA pourrait éliminer certains métiers des sciences et du génie, mais elle en transformera probablement encore davantage

Bien qu'il soit probable que l'IA perturbera le marché du travail en sciences et en génie, il est encore plus probable qu'elle transformera de nombreux métiers plutôt que de les éliminer (Lane et Saint-Martin, 2021). Brynjolfsson *et al.* (2018) ont découvert que la vaste majorité des métiers comportent à la fois des tâches extrêmement bien adaptées et mal adaptées au remplacement par l'IA et avancent que le débat sur l'impact de l'IA sur le marché du travail devrait porter sur la métamorphose des emplois plutôt que sur leur remplacement. En ce qui concerne les sciences et le génie, les conséquences pratiques de l'IA sur la façon dont la recherche est effectuée peut exiger des professionnels dans ces domaines qu'ils adoptent de nouvelles façons de penser qui reconceptualisent la valeur qu'ils apportent. Comme l'exprime un directeur de la technologie chez un fabricant de produits chimiques qui utilise l'IA pour la recherche en biotechnologie industrielle, « si vous vous voyez comme juste une personne qui règle une machine, vous pourriez bien finir par ne pas voir l'ensemble du tableau » [traduction libre] (Mullin, 2021). De plus, malgré le potentiel de l'IA de générer des hypothèses et de guider les priorités de la recherche (section 4.1), certains prétendent que les scientifiques et autres types d'innovateurs seront toujours nécessaires pour déterminer et prioriser les nouveaux problèmes et occasions (Lane et Saint-Martin, 2021).

Pour faciliter le renforcement de la capacité humaine en vue du marché du travail de l'IA, il faudra élaborer des politiques dans les domaines de l'éducation et de la formation, attirer et conserver les talents en IA et doter les gens des outils pour

utiliser efficacement les systèmes d'IA et interagir avec (OCDE, 2021b). Des pays lancent actuellement une variété d'initiatives politiques, telles que des programmes d'éducation officiels, de formation professionnelle et d'apprentissage tout au long de la vie, d'aide financière et non financière pour attirer et conserver les talents en IA et



« En ce qui concerne les sciences et le génie, les conséquences pratiques de l'IA sur la façon dont la recherche est effectuée peut exiger des professionnels dans ces domaines qu'ils adoptent de nouvelles façons de penser qui reconceptualisent la valeur qu'ils apportent. »

d'établissement de partenariats universitaires entre les établissements de recherche en IA publics et privés (OCDE, 2021b). De plus, plusieurs pays comme la France, l'Allemagne, la République tchèque et la Pologne possèdent des organismes nationaux qui suivent étroitement l'effet de l'IA sur le marché du travail. Entre 2015 et 2019, le Canada a été l'une des principales destinations des talents en IA, derrière le Luxembourg, les Émirats arabes unis et l'Irlande (OCDE, 2021b).

L'adaptation de la main-d'œuvre aux impacts de l'IA nécessitera également de réformer les programmes d'éducation et les normes et codes professionnels (Villeneuve *et al.*, 2019), d'élaborer de nouvelles normes de travail et ententes sur la main-d'œuvre et de régler les problèmes

d'EDI touchant l'IA (OCDE, 2021b). Les problèmes d'EDI exigeront de porter une attention particulière sur le manque général de diversité dans la recherche en IA et sur les inégalités dans l'accès à la technologie d'IA (section 5.4).

5.6 Impacts environnementaux des systèmes d'IA

Les systèmes d'IA peuvent avoir un gros impact environnemental en raison de l'énergie requise pour alimenter l'infrastructure de calcul. Par exemple, une analyse réalisée par Strubell *et al.* (2019) a constaté que la conception et l'apprentissage de certains modèles d'IA courants de traitement automatique des langues pouvaient produire jusqu'à 284 tonnes d'émissions de carbone, soit l'équivalent de la quantité émise en moyenne durant toute leur durée de vie par cinq automobiles. Cette estimation est toutefois contestée par Patterson *et al.* (2021), qui évaluent que les émissions de ce modèle ne sont que d'environ 5,3 % des chiffres avancés par Strubell *et al.* (2019)¹⁵.

Lacoste *et al.* (2019) ont conçu un calculateur d'empreinte carbone pour l'apprentissage machine, qui permet aux chercheurs de calculer l'impact environnemental de leurs travaux (Schmidt *et al.*, s.d.). Cependant, les

15 Pour être plus clair, Patterson *et al.* (2021) prétendent que l'estimation de Strubell *et al.* (2019) est 18,7 fois trop élevée, ce qui revient à réduire cette estimation à environ 5,3 %.

estimations ultérieures des émissions de carbone sont souvent moins précises que la mesure de l'utilisation réelle de l'énergie des systèmes d'IA (Patterson *et al.*, 2021). Des outils tels que CodeCarbon et CarbonTracker proposent du code pouvant être intégré dans les systèmes d'IA pour fournir une estimation de la quantité d'émissions de CO₂ produites (Anthony *et al.*, 2020; CodeCarbon, 2021).

Les facteurs qui influencent les émissions de carbone de l'IA incluent l'emplacement géographique et le type de l'infrastructure de calcul sur laquelle le modèle est entraîné, le type de réseau électrique utilisé et la durée de l'apprentissage (Lacoste *et al.*, 2019). Une partie de l'énergie utilisée pour alimenter l'IA peut être tirée de sources renouvelables ou compensée par l'utilisation de crédits carbone (Strubell *et al.*, 2019). Cependant, il n'est pas clair si l'achat de crédits carbone est efficace pour réduire la consommation globale d'énergie. De plus, l'énergie renouvelable n'est pas toujours accessible; au Canada, environ 60 % de l'électricité a été générée à partir de sources renouvelables (principalement l'hydroélectricité) en 2018, bien que ce ratio variait notablement selon la province ou le territoire (NRCan, 2020). Les centres de données dans le nuage optimisés pour tirer profit des énergies renouvelables sont utiles pour réduire les émissions dues au modèle d'IA (Patterson *et al.*, 2021). Patterson *et al.* (2021) proposent plusieurs exemples de techniques susceptibles d'améliorer l'efficacité énergétique des modèles d'apprentissage machine sans perte d'exactitude.

Plusieurs chercheurs (p. ex. Strubell *et al.*, 2019; Henderson *et al.*, 2020; Schwartz *et al.*, 2020; Patterson *et al.*, 2021) affirment que l'efficacité énergétique devrait être un critère d'évaluation pour la publication d'articles sur l'IA, au même rang que d'autres critères tels que l'exactitude. En outre, Gupta *et al.* (2020) insistent sur la nécessité de certifications concernant les impacts environnementaux et sociaux des systèmes d'IA afin que les consommateurs et les investisseurs puissent prendre des décisions éclairées. Taddeo *et al.* (2021) ont formulé 14 recommandations de politique au sujet d'une IA environnementalement durable. Bien que nombre de ces recommandations soient propres au contexte politique de l'UE, elles ne cherchent pas à être eurocentriques, mais à contribuer au débat international à partir d'une perspective européenne » [traduction libre].

Les impacts environnementaux de l'IA ne se limitent pas aux émissions de GES découlant du développement et de l'entraînement de modèles

Les émissions de GES produites par l'entraînement, le test et l'utilisation des modèles d'IA ne représentent qu'une petite partie des impacts environnementaux totaux de l'intelligence artificielle. Ces impacts incluent également le carbone incorporé dans les systèmes d'IA, ainsi que ceux dus à l'exploitation et à l'extraction minière et à la fabrication associées à la production des composants nécessaires

à l'IA, comme les semi-conducteurs (Mulligan et Elaluf-Calderwood, 2021). De plus, elles augmentent lorsque l'on tient compte de l'ensemble du cycle de vie des systèmes d'IA dans toute la chaîne d'approvisionnement, qui inclut l'extraction des matières premières et la fabrication des composants, le transport des matériaux et des composants, la construction et l'installation de l'infrastructure d'IA, la maintenance, la réparation, la remise en état et la mise à niveau du système et les opérations en fin de vie, comme le transport, le traitement des déchets et l'élimination (Mulligan et Elaluf-Calderwood, 2021).

Malgré ses conséquences environnementales, la recherche utilisant l'IA peut également jouer un rôle important dans la fourniture de solutions aux changements climatiques, comme la production de concepts de réseaux intelligents et d'infrastructure à faibles émissions et la modélisation de la prédiction de ces changements (Dhar, 2020). En fait, une analyse récente des programmes de financement de la recherche dans l'UE a recensé 122 projets utilisant l'IA pour s'attaquer à divers aspects des changements climatiques (Cowls *et al.*, 2021).

5.7 Sécurité de l'IA

La recherche alimentée par l'IA a probablement un vaste éventail de conséquences sur la société, dont certaines sont décrites plus haut. Cependant, le risque que cette recherche cause des préjudices sociaux pourrait être accru par la vulnérabilité des systèmes d'IA en matière de sécurité. Comme tout système informatique, l'IA est vulnérable aux cyberattaques. Ces attaques peuvent être difficiles à détecter en raison de la nature boîte noire de certains modèles d'IA (Xue *et al.*, 2020) (section 3.1.2). Les assaillants peuvent cibler les données d'apprentissage, le modèle d'IA lui-même ou l'infrastructure matérielle ou logicielle sous-jacente (AI HLEG, 2019). Ils peuvent avoir plusieurs objectifs, comme fausser ou falsifier les données d'apprentissage pour réduire l'exactitude du modèle, causer un défaut de fonctionnement du système ou lui faire faire des erreurs ou voler des modèles d'IA ou des données d'apprentissage sensibles (Xue *et al.*, 2020).

Peu de recherches examinent particulièrement le rôle de la cybersécurité pour l'IA en sciences et en génie. Néanmoins, les problèmes soulevés dans ce contexte incluent la protection de la propriété intellectuelle, la protection de la vie privée et l'exactitude scientifique. Les attaques visant à ce qu'un modèle d'IA génère des résultats inexacts pourraient avoir de graves incidences sociales négatives si ces résultats sont utilisés d'une façon qui amplifie la discrimination ou les biais sociétaux, voire nuise à la sécurité publique dans le cas des concepts de génie incorrectement évalués comme sécuritaires. Les articles 5.3 et 5.4 de l'EPTC 2 exigent des chercheurs et des établissements ou des organismes qu'ils prennent des mesures de protection couvrant l'intégralité du cycle de vie des données recueillies et utilisées dans la recherche incluant des sujets humains (IRSC *et al.*, 2018b).

Pour se défendre préventivement contre les attaques, il est nécessaire d'effectuer des évaluations de la sécurité des systèmes d'IA à l'étape de la conception, une approche que Xue *et al.* (2020) appellent *conception pour la sécurité*. Cette méthode consiste, notamment, à soumettre le système à de robustes attaques fictives, comme celles dans lesquelles on présume que l'assaillant connaît entièrement



« Les attaques visant à ce qu'un modèle d'IA génère des résultats inexacts pourraient avoir de graves incidences sociales négatives si ces résultats sont utilisés d'une façon qui amplifie la discrimination ou les biais sociétaux, voire nuise à la sécurité publique dans le cas des concepts de génie incorrectement évalués comme sécuritaires. »

le modèle, les données et les techniques de défense et a la capacité de manipuler le modèle (Xue *et al.*, 2020). Afin d'aider les développeurs en IA à assurer la sécurité de leur système, plusieurs chercheurs ont mis sur pied une bibliothèque en source ouverte d'exemples d'attaques contradictoires pouvant être utilisée pour tester la robustesse des modèles (CleverHans, 2021).

La Commission européenne (2019) et l'OCDE (2021a) citent la sécurité et la résilience aux attaques comme une exigence pour une IA digne de confiance. Cependant, contrairement à ce qui se fait ailleurs, les politiques et les stratégies d'IA du Canada ne consacrent pas beaucoup d'attention à cette question (Cussins Newman, 2019). Plusieurs pays, dont le Canada, insistent sur la nécessité de normes communes pour s'attaquer aux questions de sécurité concernant l'IA (OCDE, 2021a). Ces normes sont élaborées par des organismes tels que l'Institute of Electrical and Electronics Engineers et le British Standards Institute (Müller, 2020). De plus, le

gouvernement du Danemark, en collaboration avec des membres de l'industrie, a mis au point un mécanisme de certification, le sceau commun de cybersécurité et d'éthique des données, qui est octroyé aux entreprises qui répondent à des exigences éthiques et de cybersécurité concernant les données d'IA (OCDE, 2021b).

Les risques pour la sécurité des systèmes d'IA sont prévalents et multidimensionnels, et les parties prenantes issues des gouvernements et de l'industrie cherchent à réduire la probabilité et les répercussions d'un vol de propriété intellectuelle

Le développement de l'IA est de plus en plus vu comme une course mondiale, de nombreux pays entrant dans la danse pour exploiter cette technologie et profiter d'importants rendements économiques. Les systèmes et les outils d'IA sont déjà appliqués à des fins militaires, ce qui confère une dimension géopolitique supplémentaire à la R-D dans ce domaine (Boulanin *et al.*, 2020). L'intensification de la concurrence mondiale et l'importance des enjeux soulèvent des problèmes de sécurité.


Bien que plusieurs des risques pour la sécurité décrits plus haut dépassent le mandat du comité d'experts, certains ont des conséquences légales sur les sciences et le génie dont les gouvernements et les développeurs devraient tenir compte. Essentiellement, la précieuse propriété intellectuelle produite grâce à la R-D en IA est propice au vol ou à l'espionnage (Friesen, 2021). La nature intangible des algorithmes et des données, combinée au fait que les systèmes d'IA peuvent découler de la reconversion d'applications scientifiques à d'autres applications (Somers, 2021), a poussé plusieurs gouvernements au Canada et à l'étranger à prendre des mesures pour contrer cette vulnérabilité. En 2021, le gouvernement fédéral canadien a lancé des initiatives pour atténuer certains de ces risques en définissant de nouvelles lignes directrices pour les partenariats de recherche intégrant des questions de sécurité nationale (ISDE, 2021b) et des évaluations nationales des risques pour les chercheurs universitaires en quête de fonds fédéraux (Fife et Chase, 2021). L'IA a également été définie comme une « technologie sensible », pour laquelle les activités commerciales impliquant des investisseurs étrangers feront l'objet d'une attention soutenue dans les Lignes directrices sur l'examen relatif à la sécurité nationale des investissements récemment revues (ISDE, 2021c). Ces efforts accompagnent d'autres modifications légales, notamment aux États-Unis et dans les dispositions concernant la propriété intellectuelle des principaux traités de commerce international, qui prévoient, en particulier, des pénalités plus strictes pour le vol de secrets commerciaux (Ciuriak et Ptashkina, 2021). Les récentes réformes du *Code criminel* afin d'y intégrer les exigences convenues dans l'Accord Canada-États-Unis-Mexique (ACEUM) ont élevé le vol de secret commercial au rang d'acte criminel, ce qui traduit les craintes des gouvernements (Parlement du Canada, 2020). La législation actuelle sur la cybersécurité peut s'appliquer à ces situations et les traite comme du piratage, mais ces lois sont considérées par certains comme dépassées (Brodt *et al.*, 2021). Les autres questions légales, réglementaires et politiques concernant l'IA sont étudiées en détail au chapitre 6.

Conséquences de l'IA pour les lois, la réglementation et la politique canadiennes

- 6.1 Accès aux données
- 6.2 Commercialisation des résultats produits par la machine
- 6.3 Nouveaux risques légaux
- 6.4 Nouveaux systèmes réglementaires pour l'IA

Constatations du chapitre

- L'accès à des données de qualité est crucial pour le développement de l'IA en sciences et en génie. Les gouvernements ont édicté des politiques sur les données ouvertes, mais certaines données nécessitent une meilleure gouvernance pour protéger la vie privée, garantir une réutilisation éthique ou protéger l'intérêt de la population.
- Les données peuvent être des actifs précieux et il y a un conflit entre offrir l'accès et maintenir des avantages commerciaux. Ceci fait craindre que les données de grande valeur demeurent confidentielles.
- Obtenir la protection de la propriété intellectuelle pour plusieurs éléments constitutifs des systèmes d'IA est un défi et les découvertes ou les concepts produits sans l'humain pourraient être inadmissibles à la protection en vertu des lois actuelles.
- Les systèmes d'IA remettent en question les cadres de responsabilité juridique, en raison de leur possible manque de transparence, de l'incertitude quant à la sécurité et des défis dans l'établissement de la causalité ou dans l'attribution de la responsabilité. La résolution de ces questions varie d'un État à l'autre.
- En dépit des initiatives pour réagir l'IA dans le secteur public, la démarche réglementaire canadienne de gouvernance de l'IA nécessitera une action harmonisée entre le gouvernement fédéral et les gouvernements provinciaux et territoriaux. Les changements réglementaires apportés dans les pays partenaires peuvent également influencer le cours des réformes au Canada.

 n assiste à une course mondiale pour développer des techniques et des technologies d'IA à des fins économiques et sociétales. L'environnement juridique et politique émergent est incertain et manque d'harmonisation, ce qui crée des défis à résoudre à tous les stades du cycle de vie du développement de l'IA. À ses premiers stades, le développement de l'IA repose sur les données — leur création, leur stockage, leur acquisition leur utilisation et leur protection. À l'étape de la commercialisation, par contre, le succès requiert la gestion efficace de la propriété intellectuelle, qui n'est pas nécessairement simple avec l'IA. Ce défi touche à la fois les systèmes d'IA et leurs produits, et d'autres complications sont à prévoir à mesure que ces systèmes joueront un rôle plus actif dans la création de la propriété intellectuelle.

6.1 Accès aux données

Une caractéristique commune à tous les systèmes médias est le besoin de données. L'IA doit généralement subir un apprentissage à partir d'un énorme volume de données de grande qualité pour obtenir une précision et atténuer les biais possibles. D'un point de vue pratique, les développeurs en IA et les utilisateurs ne possèdent pas toujours toutes les données dont ils ont besoin à l'interne et cherchent alors à accéder à des sources externes. Les données utilisées pour le développement de systèmes d'IA varient considérablement sur le plan du type, de la forme et de l'origine. La variabilité des sources de données s'accompagne de



« L'importance la valeur des données pour la R-D en IA ainsi que les priorités variables entre les parties prenantes renforcent le caractère crucial de contrôler l'utilisation des données et leur circulation par des moyens légaux, des politiques et l'infrastructure »

mécanismes de gouvernance distincte. Les données personnelles sont assujetties à un certain nombre de lois provinciales ou territoriales fédérales (OPC, 2021a), alors que les données de recherche sont régies selon des politiques formelles et informelles édictées par de nombreuses parties prenantes telles que les organismes de financement et les organisations dans lesquelles se déroule la R-D.

Dans le cadre de la recherche, les systèmes d'IA tirent invariablement parti de l'énorme quantité de données disponibles en source ouverte ou par le biais de licences à partir de dépôts. Cependant, certaines parties prenantes souhaitent limiter ou restreindre l'accès à certaines données ou les protéger par des moyens légaux, tels que le droit d'auteur ou les lois sur les secrets commerciaux. L'importance la valeur des données pour la R-D en IA ainsi que les priorités variables entre les parties prenantes renforcent le

caractère crucial de contrôler l'utilisation des données et leur circulation par des moyens légaux, des politiques et l'infrastructure. Les contrats offrent aux acteurs privés une façon d'établir les modalités de l'accès aux données et de leur utilisation. Par exemple, Facebook est parvenu à inclure ses utilisateurs dans une expérience sans consentement explicite, grâce au contrat d'utilisation (encadré 3.1). D'autres organisations peuvent imposer des obligations en matière de partage des données par le biais de conditions contractuelles, qui leur permettent de rendre obligatoires dans les faits les données ouvertes.

Les gouvernements et d'autres parties prenantes cherchent à promouvoir l'IA en fournissant un accès ouvert aux données; cependant, il demeure des défis au sujet de l'interopérabilité et du respect de la vie privée

Les gouvernements sont de plus en plus conscients de la valeur pour l'innovation en IA des données qu'ils détiennent (voir The Royal Society (2017)). Pour l'UE, les données du secteur public sont un pilier de son économie numérique et en 2019, elle a créé un cadre légal pour les ensembles de données de grande valeur, auxquels on peut avoir accès de façon ouverte et qu'on peut réutiliser dans un souci de développement de l'IA (Parlement européen, 2019b). Ces dernières années, le Canada a pris des mesures pour devenir un chef de file dans les données de gouvernement ouvert (Web Foundation, 2018), avec des initiatives telles que le portail et la politique d'ouverture par défaut (GC, 2021b, 2021g). La *Charte canadienne du numérique* met également l'accent sur des cadres de données interopérables et sur un « [g]ouvernement numérique ouvert et moderne » parmi ses principes fondamentaux (ISDE, 2019b). Le gouvernement fédéral vise à fournir un accès ouvert à toutes les données de la science qu'il finance dans un format conforme aux principes FAIR d'ici à 2025 (GC, 2020c)¹⁶.

Il existe aussi des initiatives de données ouvertes dans d'autres ordres de gouvernement. Par exemple, le gouvernement du Québec a établi un partenariat avec des organismes municipaux et des organismes non gouvernementaux pour son portail de données ouvertes (Données Québec, 2021). Les gouvernements créent également de nouvelles structures et de nouveaux rôles de supervision du partage et de la gérance des données détenues par les organismes gouvernementaux et les ministères. Au Canada, le budget fédéral 2021 prévoyait des fonds pour un nouveau commissaire aux données, qui pourrait jouer un rôle dans la supervision de la disponibilité des données publiques de haute qualité et leur accès, de façon semblable à l'intendant de données en chef de la Nouvelle-Zélande ou au commissaire aux données en Australie (Gouvernement de la Nouvelle-Zélande, 2020; Gouvernement de l'Australie, 2021). À l'échelle provinciale ou territoriale, l'Ontario possède une directrice du numérique et des données depuis 2017 (Gouv. de l'Ont., 2017) et met actuellement sur pied l'Office ontarien des données pour promouvoir l'accès aux données (Gouv. de l'Ont., 2021b).

Les efforts gouvernementaux pour améliorer l'accès aux données afin que la société puisse en tirer les avantages potentiels ne se limitent pas à la diffusion, ils portent aussi sur le recueil des données et sur la mise en place de l'infrastructure de données. Cette dernière n'étant pas uniquement physique, elle est également

¹⁶ Le gouvernement fédéral a exprimé son intention d'étendre les politiques sur la science ouverte au-delà de la recherche effectuée par les organismes gouvernementaux, pour inclure la recherche qu'il finance; cependant, les détails et le calendrier de mise en œuvre étaient inconnus au moment de la rédaction du rapport (GC, 2020c).

composée d'ensembles de données et d'identifiants, de normes, de politiques, d'organismes régissant cette infrastructure et de communautés d'utilisateurs (Dodds et Wells, 2019). L'Infrastructure canadienne de données géospatiales (ICDG) constitue un exemple de plateforme complète de recueil et de partage des données spatiales (GC, 2020b). Grâce à la surveillance de caractéristiques pertinentes sur le plan géographique et des limites au Canada, l'ICDG peut éclairer une vaste gamme d'activités — de la gestion des ressources naturelles au règlement des revendications territoriales autochtones — et représente un actif public numérique polyvalent utile à de nombreuses parties prenantes et de nombreux utilisateurs (RNCan, 2019). MacGregor (2018) avance que l'expansion des activités de recueil de données de ce type — par exemple, la direction des investissements publics pour mettre à jour l'infrastructure de détection afin de se doter de capacités de communication numériques modernes — offre la possibilité de transformer les industries primaires du Canada.



« Les efforts gouvernementaux pour améliorer l'accès aux données afin que la société puisse en tirer les avantages potentiels ne se limitent pas à la diffusion, ils portent aussi sur le recueil des données et sur la mise en place de l'infrastructure de données. »

Malgré leur valeur, les données du gouvernement ouvert se heurtent à des risques et des limitations. L'augmentation de la quantité de données s'accompagne d'une hausse des exigences d'expertise et d'infrastructure pour les utiliser, une tendance qui favorise les gros acteurs institutionnels établis (Davies *et al.*, 2019). Dans le cas des données spatiales, comme dans l'exemple de l'ICDG, la fourniture et l'utilisation des données peuvent être hétérogènes

sur le plan géographique (Johnson *et al.*, 2017). Un récent rapport sur les besoins des utilisateurs de l'ICDG fait écho à ce défi et cite la nécessité d'accroître les efforts pour permettre aux communautés autochtones de tirer un meilleur parti de l'utilisation de l'Infrastructure, par exemple en offrant des ressources dans les langues autochtones (RNCan, 2019).

D'autres tensions peuvent survenir à cause du caractère sensible de certaines données. La Plateforme ontarienne des données sur la santé a été lancée dans le cadre de la réponse provinciale à la pandémie de COVID-19 et vise à offrir un meilleur accès à l'immense collection de données intégrées sur la santé (OHDP, 2020). L'utilisation de l'IA en santé dépasse le mandat du comité d'experts; cependant, il est important d'obtenir des ensembles de données sur la santé représentatifs pour améliorer la précision des modèles d'IA, comme ceux utilisés

dans les appareils médicaux (CIFAR, 2021)¹⁷. La fourniture de l'accès comporte le risque que des données sur la santé personnelle soient utilisées sans consentement ou d'une façon qui enfreigne le droit à la vie privée (Reznick *et al.*, 2020). L'accès doit être accordé conformément aux lois protégeant les renseignements sur la santé personnelle. La Plateforme ontarienne des données sur la santé exige que les chercheurs s'enregistrent et acceptent les conditions d'utilisation, dans le but d'atténuer les risques pour la vie privée qui existent lorsque des dépôts sont utilisés pour partager des données sensibles ou personnelles (Wylie, 2018; Scassa, 2019b). Des exemples comme la Plateforme, ainsi que les dispositions visant à faciliter l'accès aux données personnelles à des fins de recherche dans les propositions de réforme de la législation sur la protection de la vie privée (section 6.4), marquent la volonté des gouvernements d'accorder l'accès aux données personnelles en soutien au secteur de l'IA. Cette tendance souligne l'importance de concevoir des mécanismes légaux et de protection des données technologiques pour accompagner les initiatives de données ouvertes.

L'exploration et le moissonnage de textes et de données constituent des outils essentiels au développement de l'IA, mais ils peuvent également être illégaux et pour éliminer cet obstacle, plusieurs pays assouplissent la protection du droit d'auteur

Les renseignements tirés de l'exploration de textes et de données (ETD), comme les documents textuels, les images et les enregistrements audio, peuvent constituer une riche ressource pour l'IA. Toutefois, même lorsque les chercheurs ont la permission d'accéder aux travaux protégés par le droit d'auteur (p. ex. au moyen d'une licence d'accès à une base de données documentaire), ils peuvent ne pas avoir le droit de reproduire, de modifier ou d'utiliser ces travaux dans l'ETD à cause de la protection du droit d'auteur (Flynn *et al.*, 2020). Certains éditeurs universitaires utilisent des accords de licence pour contrôler l'accès à leurs travaux pour les demandes d'ETD (Caroll, 2019) ou aux données accompagnant les articles publiés dans leur revue (Baker *et al.*, 2019). Comme l'ETD nécessite la copie locale des données ou la conversion de documents matériels en format lisible par la machine (Flynn *et al.*, 2020), elle pourrait violer les droits économiques des détenteurs de droits d'auteur (Craig, 2021)¹⁸. Le moissonnage des données, dans lequel des outils logiciels sont utilisés pour extraire des données à partir de pages Web ou de bases de données en ligne, soulève des craintes similaires. Non seulement cette opération risque de violer le droit d'auteur par la création de

17 D'autres examens des possibilités et des risques liés à l'intégration de l'IA dans le système de santé canadien se trouvent dans Reznick *et al.* (2020).

18 Des exceptions « d'utilisation équitable » dans les lois sur le droit d'auteur permettent l'ETD quand les activités de recherche ne nuisent pas au « marché » et ne le limitent pas pour un détenteur de droit d'auteur (Craig, 2021), mais elles impliquent une analyse au cas par cas et laissent une incertitude considérable quant à la légitimité de certaines utilisations.

copies potentiellement non autorisées, mais elle peut également enfreindre les modalités d'utilisation des sites Web (Scassa, 2021a).

Dans certains pays, notamment au Royaume-Uni et dans l'UE, des exceptions explicites au droit d'auteur ont été introduites pour certaines demandes d'ETD (Kelly, 2016; Parlement européen, 2019a). Ces exceptions appuient indirectement les secteurs d'IA émergents en éliminant les risques d'infraction. Les promoteurs de ces exceptions pour l'ETD avancent qu'elles favorisent la recherche scientifique et qu'elles peuvent égaliser les chances en matière d'accès aux données tout en restant cohérentes avec les objectifs des lois sur le droit d'auteur (Craig, 2021). Leurs détracteurs, cependant, prétendent que ces exemptions empêcheront les auteurs des travaux explorés d'être justement récompensés (INDU, 2019) parce que, dans certains cas, les données contenues dans les travaux pourraient avoir exigé un investissement important pour leur création (Mercurio et Yu, 2021).

Toutes les exceptions concernant l'ETD n'ont pas la même portée, ce qui crée des possibilités de conflit entre les États. Au Royaume-Uni, les exceptions relatives à l'ETD sont limitées aux activités non commerciales (Kelly, 2016). En attendant, la directive de l'UE restreint les exceptions à des types précis d'utilisateur, tels que les organismes de recherche (Parlement européen, 2019a). Les partenariats public-privés (PPP) conclus dans le domaine de l'IA peuvent donc être exclus de ces exceptions en raison des restrictions imposées aux activités commerciales (Flynn *et al.*, 2020). Bien que l'UE tienne spécifiquement compte des PPP dans l'étendue de son cadre d'exceptions (Parlement européen, 2019a), la restriction relative à l'ETD pour les activités commerciales est vue comme une « zone sombre » au Royaume-Uni (The Royal Society, 2017). Un manque d'harmonisation semblable existe pour le moissonnage des données, l'UE et les États-Unis suivant des approches divergentes (Scassa, 2021a). Les entités canadiennes à des collaborations internationales avec des partenaires universitaires, publics ou privés devront donc se renseigner sur la légalité des éventuelles demandes d'ETD et de moissonnage qu'elles déposent, à la fois au Canada et dans les pays où se trouvent leurs partenaires. Le Canada n'a pas encore mis en œuvre les exceptions concernant le droit d'auteur pour l'ETD. Après consultation des parties prenantes, cependant, le Comité permanent de l'industrie et de la technologie de la Chambre des Communes a recommandé des modifications à la *Loi sur le droit d'auteur* afin de prévoir des exemptions explicites pour l'analyse de l'information et de permettre les activités d'ETD sans nécessiter de licence (INDU, 2019). En 2021, afin de tirer parti des données probantes collectées à partir de l'examen parlementaire, Innovation, Science et Développement économique Canada a lancé une consultation sur un cadre moderne du droit d'auteur pour l'intelligence artificielle et l'Internet des objets (ISDE, 2021a).

Obtenir la protection de la propriété intellectuelle pour les données peut être difficile, à moins qu'elles soient tenues confidentielles comme un secret commercial

La forte valeur des données pour la R-D en IA crée plusieurs conflits. D'un côté, la transparence et l'ouverture sont de plus en plus encouragées pour promouvoir la confiance et faciliter les progrès, en particulier au sein des universités et du gouvernement. De l'autre, les données sont précieuses et leur protection peut être souhaitée pour conserver un avantage commercial, bien qu'il ne soit pas toujours facile de les protéger sous forme de propriété intellectuelle. Les travaux dont les



« un des éléments constitutifs les plus précieux d'un système d'IA — les données utilisées pour le mettre au point — peut être soit inadmissible à la protection du droit d'auteur soit seulement faiblement protégé. »

données peuvent être explorées peuvent être protégés par le droit d'auteur, mais les données elles-mêmes sont en général protégées par la loi sur le droit d'auteur seulement tant qu'elles font partie d'une compilation originale (Scassa, 2019a). Et là, uniquement la sélection ou l'arrangement original des données — et pas les données sous-jacentes — est protégé (Craig, 2021)¹⁹. À ce titre, un des éléments constitutifs les plus précieux d'un système d'IA — les données utilisées pour le mettre au point — peut être soit inadmissible à la protection du droit d'auteur soit seulement faiblement protégé (Medeiros *et al.*, 2021). Bien que l'UE offre une protection *sui generis* pour les bases de données, cette protection comporte

d'importantes limitations dans le contexte de données contemporaines parce qu'il existe des exceptions pour l'extraction à des fins de recherche scientifique des données issues de bases protégées (Commission européenne, 2018).

Cette situation contribue au recours aux lois sur le secret commercial pour protéger les données. La protection du secret commercial exige que les données ou l'information demeurent secrètes, que les données aient une valeur commerciale parce qu'elles sont secrètes et que ces données fassent l'objet de mesures raisonnables, dans les circonstances, pour protéger leur caractère secret (WTO, 2017). Le maintien de la confidentialité des secrets commerciaux est essentiel à leur existence (Malone, 2020). Contrairement à la protection du droit d'auteur ou des brevets, les secrets commerciaux peuvent, en théorie, être protégés à perpétuité. La protection disparaît quand les données confidentielles tombent dans le domaine public (Scassa, 2021b). Selon Medeiros *et al.* (2021), les secrets commerciaux représentent une importante couche de la stratégie de propriété intellectuelle pour

19 Il est difficile de déterminer quand des données sont « originales », ce qui pourrait mener à l'évolution des lois sur le droit d'auteur dans ce domaine.

les entreprises technologiques d'IA émergentes, particulièrement durant la phase de développement. Au Canada, cependant, les secrets commerciaux sont généralement protégés par une loi provinciale ou territoriale, plutôt que par une loi fédérale (Scassa, 2021b). Selon la province ou le territoire, les tribunaux peuvent porter une plus grande attention aux données, ou à la façon dont elles sont recueillies ou produites, pour déterminer leur admissibilité à la protection (Malone, 2020).

Bien qu'il puisse être attrayant pour certains innovateurs de protéger leurs données par le biais des secrets commerciaux, l'efficacité de cette démarche reste à être vérifiée dans le contexte de l'IA au Canada (Malone, 2020). Scassa (2021c) et Malone (2020) affirment tous deux qu'il est possible pour l'intérêt public de surpasser la confidentialité des données dans certaines circonstances. Des mécanismes à cette fin existent tant à l'échelle fédérale qu'à l'échelle provinciale ou territoriale — comme les lois sur l'accès à l'information (GC, 1985b; MGCS, 2021) — et ils ont été employés pour obtenir le dévoilement de données d'essais cliniques détenues par Santé Canada. Cela crée des tensions ce qui concerne les engagements pris dans les accords commerciaux internationaux (Malone, 2020). Les secrets commerciaux accroissent également le manque de transparence caractéristique des systèmes d'IA boîte noire (section 3.1.1). Comme le suggèrent Citron et Pasquale (2014), en empêchant les tiers de comprendre comment et pourquoi les systèmes automatisés prennent des décisions — par exemple, dans le contexte de la détermination de la cote de crédit — la protection des données sous forme de secret commercial renforce une « société de boîtes noires », dans laquelle l'attribution des bénéfices ou des possibilités est opaque pour les personnes touchées.

Les flux de données internationaux sont un facteur croissant dans l'économie numérique mondialement interconnectée et les spécialistes de l'IA et leurs collaborateurs sont confrontés à la variabilité de la gouvernance et de la protection locales des données

Les flux de données transfrontaliers sont un problème croissant en raison de l'intégration des données numériques dans les chaînes de valeur des activités commerciales à l'échelle internationale (Lopez-Gonzalez *et al.*, 2021). Le fait que les données numériques ne connaissent pas les frontières exacerbe les défis de la gouvernance des données parce que la réglementation, la protection des données et les lois sur la propriété intellectuelle sont territoriales et donc, différentes d'un État à l'autre (Lopez-Gonzalez *et al.*, 2021)²⁰. En stipulant quelle loi régit un accord, les clauses relatives au droit applicable peuvent lever le doute à cet égard. Néanmoins, les problèmes gouvernementaux compliquent le développement des

²⁰ Considérons, par exemple, la difficulté de faire appliquer les lois canadiennes sur le respect de la vie privée contre Clearview AI pour son activité de moissonnage des images (voir l'encadré 3.2).

systèmes d'IA pour les scientifiques et les ingénieurs, qui peuvent nécessiter l'accès à des données qui ne leur sont pas accessibles dans leur pays. Par exemple, les problèmes de flux de données transfrontaliers peuvent être inévitables pour la recherche — surtout pour les collaborations de recherche — en raison de la diversité de manières dont l'ETD est abordée dans les lois nationales sur le droit d'auteur et des incompatibilités dans d'autres domaines de la gouvernance des données (Flynn *et al.*, 2020).

Les accords commerciaux modernes incluent couramment des dispositions visant à faciliter les flux de données, certains accords récents, comme l'Accord de Partenariat transpacifique global et progressiste (PTPGP) et l'ACEUM (Leblond, 2019), comportant des dispositions contraignantes et des mécanismes de mise en application (Lopez-Gonzalez *et al.*, 2021). Les ententes commerciales « mégarégionales » de ce type en sont venues à façonner le paysage des flux de données transfrontaliers, en lieu et place des accords conclus par des organismes multilatéraux tels que l'Organisation Mondiale de la Propriété Intellectuelle (OMPI), qui régit les autres formes de propriété intellectuelle (de Beer, 2020). Flynn *et al.* (2020) avancent que l'OMPI devrait néanmoins définir des normes sur les questions transfrontalières, en raison de son rôle dans la définition de la politique internationale concernant la propriété intellectuelle. Il peut exister des conflits entre les positions réglementaires et politiques internes et les exigences convenues par les responsables commerciaux canadiens et leurs partenaires internationaux. Par exemple, le PTPGP contient des dispositions non harmonisées avec les exigences de la réglementation canadienne sur la localisation des données (Leblond, 2019) et l'opérationnalisation de la *Charte canadienne du numérique* pourrait être restreinte par des dispositions de l'ACEUM (de Beer, 2020). Pour accéder à de plus grands marchés, les PME canadiennes doivent donc naviguer entre ces questions de propriété intellectuelle transfrontalières concernant les données.

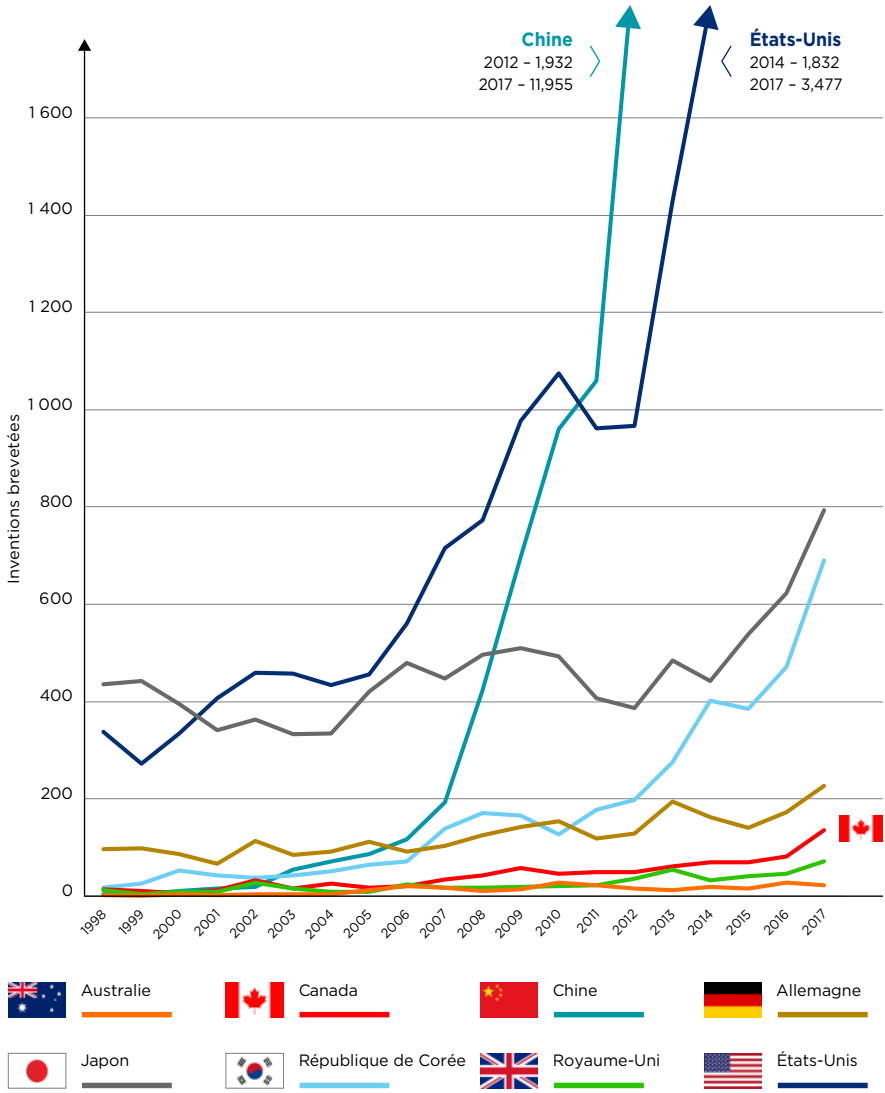
6.2 Commercialisation des résultats produits par la machine

La R-D en IA en sciences et en génie peut aboutir à des produits dont la propriété intellectuelle peut être protégée qui présentent un potentiel d'avantages économiques et sociétaux. Les résultats de ces systèmes, ainsi que leurs méthodes, leurs algorithmes, leurs données et leurs découvertes intermédiaires, peuvent représenter des formes potentiellement précieuses de propriété intellectuelle. La gestion efficace de cette propriété intellectuelle est essentielle au rendement du Canada dans le paysage de l'innovation incorporelle (Comité d'experts en matière de propriété intellectuelle, 2020; Lamb et Munro, 2020). Pour les décideurs aux échelles provinciale ou territoriale et fédérale, tirer des avantages sociétaux de la commercialisation de la propriété intellectuelle est une priorité en raison de

l'ampleur des investissements publics dans l'IA. Cette pression est vraisemblablement intensifiée par les observations persistantes — et qui ne se limitent pas à l'IA — que la propriété intellectuelle produite au Canada est mal exploitée (CAC, 2013, 2018; Hinton, 2020). Dans ce domaine extrêmement compétitif, les innovateurs canadiens devront surmonter les actuels obstacles et déterminer des approches pour résoudre les questions émergentes concernant la propriété intellectuelle produite par l'IA.

Le brevetage en IA s'est résolument accéléré ces dernières années, compliquant la situation pour les innovateurs et relançant les appels au renforcement des capacités en éducation et en gestion relatives à la propriété intellectuelle

Les brevets procurent temporairement aux innovateurs un accès protégé au marché en échange du dévoilement d'une invention. Ils sont également un important vecteur de protection et de commercialisation de la propriété intellectuelle. Les demandes de brevet doivent suffisamment dévoiler l'invention (GC, 1985a) pour qu'une personne qualifiée dans le domaine puisse la reproduire (CIPO, 2021). On assiste à une course mondiale pour breveter la propriété intellectuelle dans le domaine de l'IA, ce qui a des conséquences pour les chercheurs et les entrepreneurs canadiens (Expert Panel on Intellectual Property, 2020). La figure 6.1 révèle l'accélération du rythme du brevetage depuis la fin des années 2000, mais également des différences d'un pays à l'autre. Les chiffres — qui sont fondés sur l'emplacement du ou des détenteurs du brevet et pas nécessairement du ou des inventeurs — montrent que les plus fortes hausses de brevetage se sont produites aux États-Unis et en Chine, et plus récemment au Japon et en Corée (ISDE, 2019a).



Reproduit avec la permission du ministère de l'Industrie, 2022 (ISDE, 2019a)

Figure 6.1 Brevetage en IA dans certains pays, 1998-2017

Évolution du nombre d'inventions d'IA brevetées à l'échelle mondiale au cours des deux dernières décennies.

Les principaux secteurs de l'IA dans lesquels le brevetage a été le plus fort ont également varié et se sont diversifiés durant la même période, ce qui traduit les rapides avancées dans le domaine. De même, les principaux mots-clés et catégories de la Classification internationale des brevets ont changé au cours de la dernière décennie, « apprentissage machine » étant désormais le plus fréquent (ISDE, 2019a; Habibollahi Najaf Abadi et Pecht, 2020). Cependant, 10 % des brevets d'IA ne sont pas classifiables en vertu des sous-catégories actuelles (ISDE, 2019a). Ceci prouve à quel point la cadence de l'évolution technologique et la nature



« La nature mondiale du développement de l'IA et le fait que la nouveauté des inventions peut défier la catégorisation contribuent à la complexification du paysage de la propriété intellectuelle, dans lequel les entrepreneurs du secteur doivent s'orienter afin de fonctionner dans les marchés nationaux comme étrangers. »

générale de l'IA peuvent compliquer les efforts de suivi des tendances dans le domaine, malgré l'élaboration et le raffinement des cadres de catégorisation par l'OMPI, l'OCDE et divers autres organismes nationaux s'intéressant à la propriété intellectuelle. La nature mondiale du développement de l'IA et le fait que la nouveauté des inventions peut défier la catégorisation contribuent à la complexification du paysage de la propriété intellectuelle, dans lequel les entrepreneurs du secteur doivent s'orienter afin de fonctionner dans les marchés nationaux comme étrangers.

Cette complexité est encore renforcée par des problèmes gouvernementaux non négligeables, les exigences de brevetage pouvant varier d'un pays à l'autre (ISDE, 2019a). Par exemple, il est possible de breveter un logiciel sous forme de code informatique aux États-Unis, alors que dans l'UE et au Canada, un programme informatique doit être considéré comme une invention mise en œuvre par ordinateur pour

faire l'objet d'un brevet (OMPI, 2020; CIPO, 2021). Les problèmes gouvernementaux sont exacerbés dans certains secteurs technologiques dans lesquels les réseaux de détenteurs des droits sur la propriété intellectuelle sont particulièrement complexes, comme en biotechnologie ou en télécommunications. Cela peut conduire à des *dédales de brevets*, qui sont d'énormes grappes de brevets sur les principales technologies du domaine. Les personnes qui cherchent à innover dans ce domaine doivent alors obtenir de multiples licences, ce qui entrave considérablement l'accès et l'innovation. Les dédales de brevets incitent également au brevetage défensif, principe selon lequel des brevets sont demandés et collectés pour constituer un portefeuille qui pourra ensuite être utilisé pour gêner les entreprises concurrentes ou comme outil de négociation (Gallini et Hollis, 2019).

Pour éviter les litiges et cerner les occasions ou les stratégies de brevetage, il faut connaître les questions touchant plusieurs États et paysages de propriété intellectuelle établis. Au Canada, les innovateurs ne disposent pas toujours de la capacité ou de l'accès à l'expertise en matière de propriété intellectuelle nécessaires pour aborder ces questions (Gallini et Hollis, 2019), et la rédaction de brevets dans ce domaine exige une expertise coûteuse (Expert Panel on Intellectual Property, 2020). Résultat, seulement 2 % des PME canadiennes détenaient un brevet en 2019 (CIPO, 2019). De plus, les politiques de gouvernance de la propriété intellectuelle dans les établissements de recherche canadiens sont hétérogènes, et peuvent diverger en ce qui concerne la propriété des inventions. Un débat a lieu entre les parties prenantes quant à savoir si la diversité des approches institutionnelles procure de la souplesse aux chercheurs ou dresse des obstacles déconcertants pour les partenaires industriels (Expert Panel on Intellectual Property, 2020). Les établissements et les chercheurs individuels peuvent valoriser différemment la propriété intellectuelle et souhaiter suivre des stratégies distinctes. De récents rapports mettent en lumière des occasions de renforcer la capacité au Canada en soutenant l'éducation en matière de propriété intellectuelle (Gallini et Hollis, 2019) et en exploitant de l'expertise dans ce domaine susceptible d'être concentrée dans des zones géographiques données (Expert Panel on Intellectual Property, 2020). La Stratégie en matière de propriété intellectuelle lancée par le gouvernement fédéral en 2019 vise à procurer aux innovateurs canadiens des outils les aidant à gérer leurs actifs par le biais, entre autres, d'ateliers juridiques sur la propriété intellectuelle et de bassin de brevets (GC, 2019), mais l'efficacité de ces mesures reste à déterminer.

La protection des brevets a pour but d'encourager le dévoilement des inventions, mais l'applicabilité de cet instrument aux inventions réalisées par les systèmes d'IA est controversée

Les systèmes d'IA sont déjà capables de produire entièrement ou partiellement ce qu'on considérerait classiquement comme des résultats dont la propriété intellectuelle peut être protégée. Cependant, dans un mémoire de discussion sur la propriété intellectuelle et l'IA, l'OMPI recense des problèmes concernant la brevetabilité ou la possibilité de protéger les droits d'auteur, ce qui reflète l'absence de consensus international (OMPI, 2020). Fondamentalement, l'esprit de la protection des brevets est de définir une structure encourageant l'innovation et de faciliter le dévoilement, puis la diffusion, des inventions résultant de l'ingéniosité humaine. Une machine ne requiert pas d'encouragements pour innover; cependant, les chercheurs qui ont développé la machine peuvent réagir à des mesures incitatives visant à stimuler l'innovation permanente. Les systèmes d'IA devenant plus complexes, en particulier avec l'apprentissage machine, des

questions surgiront concernant le degré auquel une chose produite par cette machine peut ne plus être attribuée à un inventeur humain. Dans ce cas, est-ce que l'invention ne devrait plus être admissible à la protection des brevets? Pour



« Les systèmes d'IA devenant plus complexes, en particulier avec l'apprentissage machine, des questions surgiront concernant le degré auquel une chose produite par cette machine peut ne plus être attribuée à un inventeur humain »

résoudre cette question, certaines approches proposent soit d'établir un nouveau cadre juridique pour les droits sur la propriété intellectuelle dans l'IA soit de traiter la propriété intellectuelle produite par la machine à l'aide de versions modifiées des cadres existants, comme ceux concernant les inventions mises en œuvre par ordinateur (OMPI, 2020).

En vertu des lois sur les brevets, c'est l'inventeur qui dépose une demande de brevet, ce qui soulève la question de savoir si la machine peut être un « inventeur ». Selon Hagen (2021), la réponse dépend de la capacité de la machine à concevoir intentionnellement une invention en exerçant ses « processus mentaux ». Certains pays se sont montrés proactifs dans le refus du statut d'inventeur aux machines, l'Office européen des brevets (OEB) ayant résolu au début de 2020 de ne pas accorder de brevet

dans le cas d'une demande dans laquelle un système d'IA est mentionné comme principal inventeur (Hervey, 2020). D'autres ont récemment choisi d'accorder des brevets aux inventeurs machines (encadré 6.1) — d'abord en Afrique du Sud puis en Australie, cette dernière décision étant justifiée comme « cohérente avec la promotion de l'innovation » (IPWatchdog, 2021; Taylor, 2021). Les inventions produites conjointement par des humains et des systèmes demeurent brevetables; cependant, la contribution qu'un humain doit nécessairement offrir pour se qualifier comme inventeur n'est pas définie (Hervey, 2020). Il y a un risque subséquent que le statut d'inventeur ou la propriété soit réclamé dans des demandes de brevet par des humains qui n'ont pas contribué à la création de propriété intellectuelle afin de contourner les exigences de contribution et de revendiquer la protection de la propriété intellectuelle (OMPI, 2020).

Encadré 6.1 « DABUS », ou l'inventeur artificiel

Un système d'IA baptisé « DABUS » (Device for Autonomous Bootstrapping of Unified Sentience) offre un exemple de problème de brevetage concernant la propriété intellectuelle produite par l'IA. DABUS utilise l'apprentissage par renforcement pour générer des idées qui pourraient se qualifier comme invention. Les inventions que le système a produites jusqu'ici ont abouti à des demandes de brevet pour un concept de contenant à boisson et une méthode de modulation de sources lumineuses pour attirer l'attention (The Artificial Inventor Project, 2021). Des demandes de brevet nommant la machine comme inventeur ont été rejetées dans de nombreux pays, comme au Royaume-Uni, aux États-Unis, dans l'UE et à Taiwan, mais ont été acceptées en Afrique du Sud en Australie (après appel) (Chang, 2021; Egbunu, 2021)²¹. Bien que DABUS ne soit pas entraîné pour résoudre un problème particulier concernant un sujet donné (Morrison, 2019), les détracteurs de la décision de lui accorder un brevet et le statut d'inventeur citent les risques que les systèmes d'IA soient utilisés comme « producteurs de dédales de brevets » pour générer rapidement un gros volume de brevets dans des domaines ciblés afin d'écartier ses concurrents (Taylor, 2021).

L'admissibilité à la protection du droit d'auteur et son caractère approprié pour des créations générées par la machine sont incertains en raison d'incompatibilités avec le cadre juridique actuel

Un important débat concernant les lois sur le droit d'auteur cherche à décider si les travaux produits par la machine devraient être protégés et si oui, comment en déterminer la paternité. Au Canada, les travaux produits par des « auteurs » sont protégés par le droit d'auteur. Craig (2021) suggère que, bien que la *Loi sur le droit d'auteur* ne définisse pas explicitement les auteurs comme des humains, elle rend cette définition implicite par ses exigences. Par exemple, les auteurs admissibles doivent être résidents d'un emplacement géographique donné et la durée du droit d'auteur est liée à la durée de vie de l'auteur. La nature anthropocentrique de la législation sur le droit d'auteur pose d'autres problèmes lorsqu'il s'agit de déterminer l'admissibilité à la protection du contenu. Le seuil subjectif concernant « l'originalité » requise pour qu'une création soit admissible à la protection du droit d'auteur n'est pas facile à appliquer aux créations de machines (Craig, 2021).

²¹ Les demandeurs ont fait appel de décisions négatives. Cependant, les tribunaux du Royaume-Uni ont rejeté l'appel concernant leur décision de ne pas donner suite à la demande (Bond, 2021).

Le caractère approprié du droit d'auteur pour des travaux artistiques produits par la machine fait actuellement l'objet de débats (Deltorn, 2017; Craig et Kerr, 2021) et il est possible de faire des parallèles avec les créations réalisées en sciences et en génie. Il n'est pas certain si la protection du droit d'auteur au Canada sera réformée ou non, mais le gouvernement fédéral a entamé une consultation publique sur la modernisation de son cadre de droit d'auteur avec l'IA explicitement en tête (GC, 2021a). Craig (2021) avance que si l'on se fie à la *Loi sur le droit d'auteur* canadienne et à l'histoire du droit d'auteur, l'absence de protection du droit d'auteur pour les créations générées par la machine n'entraînerait pas de « sous-production » de travaux originaux, parce que cette protection n'encourage pas les machines. Dans le contexte canadien, on pourrait vraisemblablement établir un processus autonome encourageant la production d'artefacts par la machine afin d'offrir une protection limitée analogue au droit d'auteur (droits, modalités, monopoles limités), mais ce processus devrait être neutre technologiquement pour éviter de devenir rapidement désuet (Craig, 2021). Ou alors, les résultats d'IA atteignant le seuil « d'originalité » pour être admissibles au droit d'auteur pourraient simplement être du domaine public par défaut (Craig, 2021). La vitesse à laquelle le contenu généré par la machine peut être créé est si grande que, si ce contenu était protégé, il serait difficile pour les créateurs humains d'entrer en compétition et d'éviter des infractions (Asay, 2020). Conserver les résultats produits par la machine dans le domaine public élimine ce problème et permet plutôt au contenu d'inspirer les créateurs humains (Craig, 2021).

L'incertitude concernant la possibilité d'obtenir un brevet pour des inventions de machines peut entraîner le recours aux lois sur les secrets commerciaux, qui réduit la transparence

En dépit des indices de prochaines réformes du droit d'auteur au Canada, les dispositions d'adaptation des cadres de propriété intellectuelle afin de répondre aux défis posés par l'IA n'ont pas encore été implantées; cependant, les problèmes soulevés dans la section précédente dépeignent un environnement technologique mondial en rapide évolution dont les lois et les politiques ont du mal à suivre le rythme. À l'heure actuelle, des exceptions à la protection du droit d'auteur ont été proposées ou mises en œuvre de façon hétérogène dans certains pays pour promouvoir le développement de l'IA. En même temps, la possibilité de brevetage des éléments constitutifs des systèmes d'IA varie d'un pays à l'autre. Même à l'intérieur d'un même pays, des affaires actuellement devant les tribunaux influenceront sur la probabilité qu'une invention fasse l'objet d'un brevet. Une analyse des décisions du United States Patent and Trademark Office entre 2013 et 2020 dans différents domaines dans lesquels l'IA est appliquée a constaté que le taux de rejet des demandes de brevet variait selon le secteur d'activité et que les décisions des tribunaux pour certaines causes ayant fait date avaient eu des conséquences considérables sur les décisions ultérieures (Gaudry et Vandsburger, 2020). Par

conséquent, il existe une incertitude quant à l'efficacité des brevets et du droit d'auteur pour protéger la propriété intellectuelle de l'IA. Ce problème sera encore complexifié par les questions d'admissibilité décrites plus haut dans les situations où les humains sont absents de la boucle de conception et de découverte.

On s'inquiète qu'une protection de la propriété intellectuelle moins formelle réduise la transparence et la confiance dans les systèmes d'IA et appauvrisse le domaine public (OMPI, 2020; Hagen, 2021). Dans l'économie incorporelle, les secrets commerciaux peuvent être attrayants, non seulement en raison de l'incertitude entourant la possibilité d'obtenir un brevet pour des inventions générées par la machine, mais également parce que la protection des secrets commerciaux a été renforcée dans certains pays (Ciuriak et Ptashkina, 2021). Bien que la robustesse et l'admissibilité de la protection des secrets commerciaux pour les algorithmes et les données soient aussi entourées d'une certaine incertitude (Hagen, 2021), les avantages concurrentiels peuvent être maintenus tant que ces secrets demeurent. La confidentialité peut être facilement maintenue, surtout quand il est difficile de soumettre les systèmes d'IA à la rétro-ingénierie (OMPI, 2020)²².

Plusieurs conséquences problématiques pourraient survenir si les innovateurs en IA adoptent les secrets commerciaux comme principal outil de protection de la propriété intellectuelle. Premièrement, les secrets commerciaux peuvent entraîner la diminution du dévoilement public des découvertes (Hagen, 2021), ainsi que des « secrets commerciaux négatifs », soit les secrets commerciaux concernant l'information sur des inventions ou des approches qui ne fonctionnent pas ou qui sont un échec. Cette information peut être aussi précieuse que la connaissance des approches prometteuses ou fructueuses parce qu'elle évite de se retrouver dans une impasse en R-D (Pritt, 2018). Deuxièmement, on craint que les acteurs établis dans le marché de l'IA bénéficient de façon disproportionnée de la confidentialité et qu'il y ait des raisons sociales ou éthiques pour suspendre la protection des secrets commerciaux afin de ne pas exacerber les lacunes technologiques préexistantes à l'échelle mondiale (OMPI, 2020). Enfin, la protection des secrets commerciaux peut aussi avoir un effet contraire à plusieurs exigences réglementaires émergentes, telles que le « droit à l'explication » dans le système de prise de décision automatisée (PDA) (Hagen, 2021).

De plus, la protection des secrets commerciaux est fragilisée par la forte mobilité de la main-d'œuvre dans le domaine de l'IA et par le paradigme de l'innovation ouverte (Ciuriak et Ptashkina, 2021). Les plateformes logicielles en source ouverte sont populaires dans la mise en œuvre de l'apprentissage machine. Beaucoup de ces plateformes — mises au point par des géants issus

²² Il est possible d'extraire des modèles d'apprentissage machine par le vol de modèle (section 5.7), les puissants outils d'IA exclusifs offerts contre des frais d'utilisation représentant des cibles potentielles (Brod et al., 2021).

de la recherche scientifique — sont parmi les outils les plus avancés disponibles gratuitement (Engler, 2021). Ainsi, la valeur réelle la propriété intellectuelle reposera sur des combinaisons de logiciels et de données. Même si les logiciels et les données sont techniquement ouverts, les données les plus utiles, ou la combinaison ou les moyens pour combiner les logiciels et les données, peuvent être cachés sous forme de secret commercial. Toutefois, certaines solutions réalisent l'équilibre entre le désir des spécialistes de maintenir la confidentialité et l'obtention de la transparence nécessaire. Par exemple, les organismes de réglementation peuvent agir comme des intermédiaires tiers quand le dévoilement est nécessaire pour les besoins de la reddition de comptes (comme dans le contexte de la PDA), mais que la confidentialité est requise parce que l'information est considérée comme un secret commercial (Hagen, 2021). Néanmoins, les

décideurs sont mis au défi de définir des mesures incitatives et des mécanismes de reddition de comptes qui promeuvent le développement et évitent les effets problématiques décrits ci-dessus.

Les collaborations en IA sont fréquentes et réunissent des parties prenantes dont les priorités et les motivations sont variées; les ententes précoces et proactives pour clarifier les droits de propriété intellectuelle entre les collaborateurs peuvent alors favoriser la commercialisation et la rétention de la propriété intellectuelle au Canada

Comme il est noté aux sections 2.1 et 6.1, il existe une grande interdisciplinarité dans la R-D en IA (The Royal Society, 2017) et les collaborations sont fréquentes tant au Canada qu'à l'étranger (Wu *et al.*, 2020). Par définition, la Stratégie pancanadienne en matière d'intelligence artificielle du CIFAR a favorisé les nouvelles connexions dans le réseau de recherche en IA grâce à ses programmes de financement, encourageant l'accroissement des investissements étrangers (Chowdhury *et al.*, 2020) et la création de gros PPP avec des chefs de file multinationaux en IA (Brandusescu, 2021). Bien que ces partenariats soient essentiels pour créer une masse critique de R-D en IA au Canada, certains avancent que la présence de gros acteurs étrangers dans le domaine présente des risques qui pourraient miner les approches « fait au Canada » (Brandusescu, 2021). La propriété intellectuelle et les données sont extrêmement précieuses pour ces acteurs, qui possèdent une capacité à exploiter ces actifs difficile à égaler au pays

(Gallini et Hollis, 2019). Les PPP font fréquemment en sorte que les droits sur la propriété intellectuelle développée au Canada appartiennent à des étrangers et que des entreprises en démarrage canadiennes sont acquises par des entreprises étrangères (Hinton et Cowan, 2018). En raison de cette situation, les PME qui mènent des activités au Canada peuvent avoir de la difficulté à obtenir des licences, et être alors forcées de payer des royalties pour des travaux protégés en vertu de la propriété intellectuelle réalisés au pays, voire tout bonnement ne pas obtenir de licence (Hinton, 2020). Les craintes concernant le degré auquel la recherche financée publiquement se traduit par une propriété intellectuelle qui, finalement, quitte le Canada ne se limitent pas à l'IA, mais c'est dans ce domaine que le problème est le plus notable. En date de 2020, des entreprises étrangères avaient acquis la majorité des brevets d'IA et d'apprentissage machine obtenus à la suite de demandes déposées au United States Patent and Trademark Office par des chercheurs et des établissements situés au Canada depuis 2015 (Hinton, 2020)²³.

Medeiros *et al.* (2021) souligne l'importance des contrats pour surmonter les problèmes de droits et de propriété concernant les avancées en IA réalisées au Canada, mais également pour éviter les futurs conflits entre les partenaires. Les contrats — y compris les accords de licence, commerciaux et de collaboration — peuvent jouer un rôle important dans la structuration des relations ou dans la gouvernance de l'accès aux données ou à la technologie et leur utilisation. Par exemple, il est possible d'utiliser des ententes de collaboration pour préciser l'apport de chaque membre dès le début d'une collaboration (p. ex. éléments matériels et logiciels). Ces ententes peuvent également définir les droits concernant les travaux dont la propriété intellectuelle peut être protégée ultérieurement. Cette mesure de protection peut garantir qu'une partie d'un projet ne cède pas le contrôle technologique à une autre, par exemple, quand elle partage son expertise (Medeiros *et al.*, 2021). Les partenaires pourraient utiliser des contrats pour délimiter l'utilisation prévue des technologies développées durant le partenariat (p. ex. utilisation commerciale ou non commerciale). Dans le cas d'applications non prévues préalablement, cette définition peut être bénéfique aux petites entités qui souhaitent éviter les litiges concernant à ces utilisations non intentionnelles ou être tenues responsables de celles-ci.

Il est également possible de mieux protéger les intérêts et la propriété intellectuelle des parties prenantes non commerciales ou des petites entités à l'étape de l'attribution de licence. Par exemple, comme ces acteurs peuvent ne pas disposer des ressources suffisantes pour s'engager dans des litiges étendus avec

23 Certains établissements canadiens de recherche en IA ont édicté des mesures concernant la propriété intellectuelle dont leurs collaborateurs potentiels devraient tenir compte. La politique du Vector Institute décrit la commercialisation de sa propriété intellectuelle d'une façon qui « met l'accent sur le développement économique par la création d'entreprises en démarrage canadiennes » [traduction libre] (Vector Institute, s.d.), alors que Mila met de l'avant une politique de dévoilement public de la propriété intellectuelle par la publication plutôt que par le brevetage (Mila, 2021).

des entreprises multinationales, les ententes de collaboration sous-tendant les gros PPP pourraient stipuler que les PME établies au Canada sont autorisées à obtenir des licences sur les droits à la propriété intellectuelle générée par le partenariat, peut-être sans avoir à verser de royalties pendant une période limitée (Bawa et Tawfik, 2020). Ce modèle a été employé dans la mise au point des vaccins contre la COVID-19 effectuée au sein d'un PPP au Royaume-Uni. La participation accrue dans des communautés de brevets — comme celles créées dans le cadre de la Stratégie pancanadienne en matière d'intelligence artificielle — et une meilleure exploitation de l'expertise en propriété intellectuelle que l'on trouve dans les supergrappes pourraient également faciliter une utilisation plus efficace de la propriété intellectuelle développée au Canada par le biais de partenariats avec des entreprises multinationales étrangères (Bawa et Tawfik, 2020).

6.3 Nouveaux risques légaux

Le déploiement d'un système d'IA peut élargir ses interactions avec la société. Bien que les scientifiques, les ingénieurs et les scientifiques des données (entre autres) puissent passer des mois ou des années à tester et à entraîner un tel système, l'application de ce système à un nouvel environnement peut avoir des conséquences imprévues. De nombreux experts participent au développement et au déploiement des systèmes d'IA, à un point tel qu'il pourrait être impossible de déterminer une « personne physique »²⁴ unique responsable du système ou de ses actions. Les circonstances sont également à considérer : les systèmes destinés à la conception et à la découverte en sciences et en génie peuvent être spécialisés et ne pas être aussi sujets aux résultats problématiques de façon extrêmement publique (Schwartz, 2019). Cependant, ces systèmes seront probablement intégrés dans leur environnement et les spécialistes doivent tenir compte du potentiel d'infractions et de préjudices, ainsi que de la nécessité de maintenir des systèmes sécurisés étant donné leur aptitude à informer ou à assumer des fonctions décisionnelles.

L'opacité de certains systèmes d'IA entrave l'application des principes de responsabilité juridique, de façon que l'application des droits de propriété intellectuelle et la protection de la vie privée ne sont plus simples

Les systèmes d'IA nécessitent l'accès à de vastes quantités de données et, dans certains cas, réécrivent des segments de code dans les algorithmes responsables de leur fonctionnement. Dans de telles situations, le système fait face au risque d'enfreindre des droits de propriété intellectuelle lorsqu'il accède à des ensembles de données protégés par le droit d'auteur, les copie ou les manipule, ou quand il

24 Une « personne physique » est définie comme un être humain, par opposition à une « personne morale » ou une « personne juridique », qui peut être un organisme privé ou public.

implémente des processus protégés par brevet (Medeiros *et al.*, 2021).

Traditionnellement, les détenteurs de droits de propriété intellectuelle peuvent poursuivre les entités responsables d'infractions pour faire valoir leurs droits. Cependant, il est alors crucial de justifier qui est responsable de l'infraction, comment la responsabilité est divisée entre de multiples parties, s'il y a lieu, et quelles actions exécutées par le système ont causé l'infraction (Benhamou et Ferland, 2021).

Comme il est noté plus haut, la création d'un système ou d'un outil d'IA fait généralement appel à de multiples acteurs, et il pourrait être impossible d'attribuer la responsabilité à un seul individu pendant la durée du développement (Giuffrida, 2019). De plus, l'IA et son écosystème de données (ainsi que les appareils qui créent et recueillent ces données) ne sont pas séparables, ce qui complique encore l'attribution de la responsabilité (Giuffrida, 2019). Enfin, si le système d'IA est une boîte noire, il peut être impossible d'établir la nécessaire relation de cause à effet qui relie les actions du système à l'infraction. Le manque de transparence peut empêcher la détection de l'infraction au droit d'auteur durant la phase d'apprentissage, encore moins l'attribution de ces actions à des individus (Craig, 2021). La présence ou non d'un humain dans la boucle et le fait que l'IA soit considérée comme un produit ou comme un service (ou comme intégrée dans un appareil) dicteront la manière dont la responsabilité pourrait être attribuée de façon prévisible (Giuffrida, 2019).

Bien que certaines activités de recherche puissent être protégées du risque de litige²⁵, il demeure des inquiétudes quant au potentiel d'imprévisibilité dans l'application de la propriété intellectuelle (Medeiros *et al.*, 2021). Medeiros *et al.* (2021) prétendent qu'en l'absence de nouvelles mesures législatives telles que les exemptions concernant l'ETD dans la *Loi sur le droit d'auteur* (section 6.1), les ententes contractuelles peuvent aider à combler les fossés actuels. Les contrats et les accords de licence entre des entités juridiques peuvent définir les droits et les obligations relatifs à l'utilisation de travaux protégés en vertu du droit d'auteur par des systèmes d'IA. Dans un même temps, il est concevable que les détenteurs de licence et les utilisateurs puissent profiter d'une exonération de certaines obligations concernant la propriété intellectuelle, laissant les propriétaires du produit régler les revendications à ce sujet (Medeiros *et al.*, 2021). Bien que les contrats ne soient pas une solution aux problèmes étendus concernant la responsabilité liée aux infractions touchant la propriété intellectuelle, cet outil peut être utile aux détenteurs de propriété intellectuelle à titre de mécanisme de protection et de reddition de comptes.

25 Par exemple, les activités à l'intérieur des limites d'utilisation équitable du droit d'auteur, qui permettent cette utilisation de matériel protégé par ce droit (Craig, 2021).

Les législateurs et les juristes se penchent sur des cadres de responsabilité et de reddition de comptes qui équilibrent le désir d'aider le développement de système d'IA et la nécessité de protéger contre les préjudices

Il est vraisemblable que les systèmes d'IA puissent causer d'importants préjudices à l'origine de poursuites judiciaires. Par exemple, les systèmes d'IA qui surveillent ou entretiennent l'infrastructure (Sparkes, 2021) et ceux qui créent de nouvelles formes de vie artificielle susceptibles de s'intégrer à un écosystème naturel (Kriegman *et al.*, 2020) peuvent présenter des risques de préjudice consécutif à la mise en pratique de l'IA. On s'attend à ce que la réglementation légale protège les gens contre ce genre de préjudice, tout en encourageant l'innovation. Comme la responsabilité civile déterminée par les tribunaux en vertu de l'application des principes juridiques généraux aux nouveaux scénarios introduits par les systèmes



« La question est de savoir si la réglementation de l'IA nécessite l'adaptation des cadres légaux actuels ou l'élaboration de nouveaux cadres. »

d'IA peut ne pas être le moyen le plus efficace ou efficace pour compenser le préjudice causé par l'IA, une réforme législative est inévitable (Giuffrida, 2019). La question est de savoir si la réglementation de l'IA nécessite l'adaptation des cadres légaux actuels ou l'élaboration de nouveaux cadres (Giuffrida, 2019).

Les principes de responsabilité diffèrent selon qu'ils sont appliqués à des produits ou à des services. La détermination du régime de responsabilité qui s'applique aux systèmes d'IA dépend de plusieurs facteurs, dont la question de savoir si l'IA constitue

ou non une « forme matérielle » (Chagal-Feferkorn, 2019). Bien que le concept « d'autonomie » soit largement utilisé pour faire la distinction entre les technologies traditionnelles et sophistiquées, il n'est pas nécessairement utile pour décider quel régime de responsabilité doit s'appliquer (Chagal-Feferkorn, 2019). La responsabilité des produits ne régit pas toujours adéquatement l'IA pour plusieurs raisons. Par exemple, l'IA peut causer des préjudices en l'absence de défaut dans son système, un élément nécessaire à la responsabilité des produits (Benhamou et Ferland, 2021). De plus, le problème de transparence créée par les systèmes d'IA boîte noire limite la prévisibilité des résultats de ces systèmes, un autre élément requis de la responsabilité des produits (Benhamou et Ferland, 2021).

Plusieurs solutions ont été proposées pour faire face aux difficultés à concilier les cadres classiques de responsabilité et les processus et résultats accompagnant les systèmes d'IA pouvant posséder divers degrés d'autonomie (Benhamou et Ferland, 2021). Les régimes actuels de responsabilité des produits et des services pourraient être étendus afin de tenir les producteurs ou les exploitants des systèmes d'IA

automatiquement responsables par le biais de divers mécanismes (Benhamou et Ferland, 2021). On pourrait également accorder le statut de personne morale à l'IA, ce qui résoudrait certains des problèmes constatés, mais pourrait également avoir d'importantes conséquences sur les lois relatives au droit d'auteur et aux brevets (section 6.2). Toutefois, accorder le statut de personne morale à l'IA apporterait son lot de défis, étant donné la difficulté précédemment mentionnée d'identifier une personne physique responsable (Benhamou et Ferland, 2021).

Il n'y a actuellement pas de consensus international au sujet de la voie à suivre, mais la réglementation de l'IA proposée par l'UE donne une idée de la façon dont elle a l'intention de procéder. La personnalité morale n'est pas actuellement vue comme une solution viable, pas plus que la révision du régime de responsabilité existant dans l'UE et dans ses pays membres (Parlement européen, 2020). Le Parlement européen envisage plutôt la révision des cadres de responsabilité des produits, ainsi que la « responsabilité stricte » pour les systèmes d'IA autonomes qui sont jugés à risque (Parlement européen, 2020; Commission européenne, 2021a). Dans ce cadre, les exploitants des systèmes d'IA seraient tenus responsables des préjudices, même s'il n'est pas démontré qu'ils sont en faute, en raison des risques inhérents à l'incontrôlabilité et à l'imprévisibilité du système qu'ils utilisent (Parlement européen, 2020). Giuffrida (2019) affirme que les approches de la responsabilité fondées sur le risque ou sur les préjudices possèdent des parallèles avec les approches actuelles de réglementation de la vie privée et de la cybersécurité et offriraient un moyen simple de garantir la compensation des préjudices par le biais d'une certaine forme d'assurance-responsabilité (Parlement européen, 2020). La démarche suivie par l'UE converge vers une approche harmonisée et reconnaît le risque inhérent au vide réglementaire entourant la responsabilité du point de vue des droits humains et du développement économique (Parlement européen, 2020).

6.4 Nouveaux systèmes réglementaires pour l'IA

Comme nous l'avons examiné, des conflits sont perceptibles entre la réglementation de l'IA et l'encouragement d'un environnement propice à l'innovation en IA dans d'autres secteurs. De nombreuses parties prenantes sont concernées, chacune jouant un rôle distinct : les décideurs peuvent fournir des orientations quant à ce qu'ils considèrent qui *devrait* être fait pour réglementer l'IA, mais ce qui *peut* être fait dépendra fortement de l'état du droit. Les interactions entre les décideurs et les législateurs, d'une part, et les développeurs et les utilisateurs de l'IA externes, d'autre part, présentent de nombreux points de friction. Jusqu'ici, les interventions réglementaires concernant l'IA — en particulier les interventions préventives — sont vues comme soutenant l'innovation en contribuant à la confiance dans l'IA par de multiples parties prenantes (Deloitte Canada, 2019; LCO, 2021). Ces interventions peuvent aussi atténuer l'incertitude juridique entourant le déploiement des

systèmes d'IA, qui pourrait dissuader l'investissement (Giuffrida, 2019). Par exemple, l'UE offre un vaste « marché numérique unique » par le biais de nombreuses lois et politiques, principalement par le RGPD, mais aussi grâce à sa récente proposition de cadre réglementaire exhaustif de l'IA (Mercurio et Yu, 2021).

Plusieurs démarches canadiennes de développement et de déploiement responsables de l'IA et de ses technologies ont été élaborées à ce jour dans un paysage complexe et incomplet, et d'autres sont en développement (LCO, 2021). Cette complexité est en partie due à la division des pouvoirs prévue par la constitution canadienne et fait en sorte que divers volets de la réglementation de l'IA sont répartis entre le fédéral et les provinces et territoires. Les gouvernements provinciaux et territoriaux disposent d'une considérable autonomie dans plusieurs domaines intéressant le développement de l'IA, comme la santé et l'éducation. Les compétences relatives à la protection de la vie privée et des données sont également fragmentées entre les secteurs public, privé et celui de la santé. L'environnement réglementaire canadien de l'IA commence à prendre forme autour de la conception de plusieurs politiques qui s'effectue de manière asynchrone dans les différents ordres de gouvernement. Toutefois, étant donné la dimension internationale de la gouvernance et du développement de l'IA (sections 6.1 et 6.2), les décisions prises localement au Canada peuvent avoir des conséquences mondiales et les décisions mondiales, des conséquences locales. Les environnements réglementaires refléteront les différentes valeurs et normes, ce qui aura des conséquences sur la collaboration et le commerce transfrontaliers et pourrait aboutir à des incompatibilités entre chaque côté des frontières.

Bien que le gouvernement du Canada ait implanté un cadre de gouvernance de l'IA, la portée de ce cadre est limitée à la fonction publique fédérale, ce qui crée un vide réglementaire

Au Canada, l'initiative Utilisation responsable de l'intelligence artificielle a établi un cadre de gouvernance de l'IA au Canada à l'échelle fédérale (GC, 2021f). Ce cadre est composé d'un ensemble de principes directeurs, d'une liste de fournisseurs qualifiés pour fournir des services aux ministères et organismes gouvernementaux au moyen de systèmes d'IA, d'un outil d'évaluation de l'incidence algorithmique et de la Directive sur la prise de décisions automatisée (GC, 2021f). La Directive est le pivot du cadre en raison de la capacité des systèmes d'IA à influencer négativement sur les droits humains et à éroder la confiance, et il a été loué pour sa nature préventive (LCO, 2021). Cependant, sa portée est intrinsèquement limitée : il ne s'étend pas aux gouvernements provinciaux et territoriaux ou aux administrations municipales en raison de la division des pouvoirs au Canada. Il ne s'applique pas non plus aux activités du secteur privé, à moins qu'elles concernent des produits ou des services achetés par le gouvernement fédéral (LCO, 2021). Il en résulte un vide réglementaire

(LCO, 2021). Malgré les efforts actuellement déployés par certains gouvernements provinciaux ou territoriaux pour mettre en place des cadres équivalents (Gouv. de l'Ont., 2021c), de nombreuses applications possibles de la PDA (et d'autres utilisations de l'IA) par les organismes provinciaux, territoriaux et municipaux ne sont pour l'instant pas réglementées²⁶. Plusieurs ont critiqué l'environnement réglementaire actuel de l'IA au Canada pour son « approche ponctuelle » et le manque de cohérence entre les différents ordres du gouvernement (Brandusescu, 2021; McKelvey et Roberge, 2021).

Les lois canadiennes régissant l'accès à certains types de données personnelles contribuent aussi à la réglementation de l'IA, quoiqu'indirectement, dans un secteur qui est également compliqué par la division des pouvoirs. Les lois sur le respect de la vie privée visant à régir l'utilisation des données personnelles sont fragmentées selon le type d'organisme (secteur public ou privé) qui utilise ces données, l'endroit où les données sont situées, le type de données et si les données traversent ou non les frontières provinciales ou territoriales (OPC, 2021a). L'absence d'harmonisation contribue à rendre difficiles la protection et le partage des données. Des tentatives sont en cours pour réformer et moderniser la législation afin de tenir compte des possibilités et des risques présentés par l'IA et les mégadonnées, particulièrement pour les activités commerciales. À l'échelle fédérale, le projet de loi C-11 (*Loi de 2020 sur la mise en œuvre de la Charte du numérique*) visait à réviser la législation fédérale concernant le respect de la vie privée applicable aux données personnelles utilisées par le secteur privé (Cofone et al., 2021), mais il a été abandonné avec le déclenchement des élections fédérales à l'automne 2021 (Smith, 2021). De son côté, le Québec a récemment adopté le projet de loi 64 (Assemblée Nationale, 2020) dans le but de combler les lacunes de la précédente loi, qui avait été jugée « mésadaptée au nouveau contexte généré par le développement rapide et l'adoption massive des technologies numériques » (CEST, 2021). Les secteurs dans lesquels la précédente loi était vue comme inadéquate comprenaient le traitement des renseignements personnels susceptibles d'être inférés par les systèmes d'IA, les moyens par lesquels les données pouvaient être utilisées pour la recherche et l'absence d'une disposition sur les droits individuels concernant la prise de décision automatisée. Il est à noter que l'Ontario a proposé sa propre loi de protection des données du secteur privé, qui reprendra des dispositions à la fois du projet de loi C-11 fédéral et du projet de loi 64 québécois (Gouv. de l'Ont., 2021a).

Les utilisations des systèmes d'IA en sciences et en génie pour la conception et la découverte pourraient ne pas constamment recourir aux données personnelles et donc ne pas être concernées par ces réformes. Néanmoins, le comité d'experts

²⁶ Les États-Unis, le Royaume-Uni et la Nouvelle-Zélande utilisent actuellement des outils de PDA pour faciliter de nombreuses opérations dans les champs de la justice pénale, administrative et civile (LCO, 2021).

pense que les réformes proposées aux lois régissant les données du secteur privé pourraient servir de pivot à un futur cadre réglementaire plus vaste.

Les défis de la réglementation de l'IA incitent à utiliser de nouvelles approches réglementaires qui seront soumises à un examen et à un renouvellement permanents

Il existe de nombreux scénarios dans lesquels les outils légaux et réglementaires se révèlent inadéquats pour corriger les situations générées par le déploiement de systèmes d'IA en dehors de la recherche en laboratoire. Plusieurs pays ont commencé à réformer ou à adapter leurs lois et leurs processus pour se préparer à des systèmes d'IA de plus en plus présents et autonomes, mais certains explorent également de nouveaux outils réglementaires qui pourraient être plus adaptables et souples face à la rapide évolution de la technologie. Plusieurs expérimentations sont effectuées sous la forme de bacs à sable réglementaires, qui sont conçus pour faciliter aux innovateurs les tests pratiques en leur offrant un environnement contrôlé, de façon que les apprentissages puissent fournir aux organismes de réglementation des données pour l'élaboration de politiques. Les bacs à sable peuvent prendre de nombreuses formes, des dispenses réglementaires limitées dans le temps à des emplacements physiques précis dans lesquels l'expérimentation peut se dérouler (OCDE, 2021b). L'échelle de ces derniers varie et ils ont été fixés dans des établissements et des laboratoires de recherche, le long de routes ou dans des régions géographiques complètes²⁷. Plusieurs pays ont soit développé des bacs à sable soit ont l'intention de le faire en fonction de leur stratégie nationale (Kung *et al.*, 2020). Un bac à sable réglementaire a déjà été introduit dans le système de santé canadien pour les produits thérapeutiques fondés sur les nouvelles technologies, y compris sur l'IA (SC, 2019). Les détails de sa mise en œuvre ne sont pas encore entièrement connus (SC, 2021), mais certaines caractéristiques du nouveau mécanisme ont déjà été appliquées face à la pandémie de COVID-19 dans les exigences de conception d'essais cliniques et de production de rapports (Eren Vural *et al.*, 2021). Un des éléments essentiels est la quantité de données probantes devant être présentées aux organismes de réglementation et le temps à attendre avant qu'un produit puisse commencer à être utilisé par les patients (SC, 2019). Des stratégies semblables pourraient être appliquées aux cadres réglementaires concernant les systèmes d'IA pour, potentiellement, accélérer la mise en marché; cependant, elles obligerait les gouvernements à gérer les risques et à étudier à la loupe les résultats après la mise en marché afin de guider les ajustements nécessaires à la réglementation.

²⁷ La Chine a établi une « zone pilote » pour le développement de l'IA dans un comté de 500 000 habitants, dans laquelle une réglementation flexible permettra l'expérimentation de véhicules autonomes, de l'agriculture intelligente et de l'IA au sein du gouvernement (Xinhuanet, 2020).

Des examens et un renouvellement réguliers de la politique actuelle sur à l'IA peuvent aider à éviter les résultats négatifs et à tirer profit des nouvelles possibilités. Dans son principe 2.3, l'OCDE recommande que le déploiement des systèmes d'IA de confiance inclue la détermination des moyens « d'améliorer l'adaptabilité, la réactivité, la polyvalence et l'applicabilité des instruments politiques » qui s'appliquent à l'IA (OCDE, 2021d). La prévalence des approches d'autoréglementation jusqu'ici ne fait que souligner le besoin de transparence entre les développeurs et les autorités réglementaires afin que des examens et des évaluations puissent avoir lieu (Renda, 2019). Au Canada, la Directive sur la prise de décisions automatisée est assujettie à un examen par les pairs effectué tous les six mois, qui regroupe des parties prenantes du secteur public et de l'extérieur (GC, 2021f). De son côté, la Commission européenne encourage ses États membres à examiner et à mettre à jour leur stratégie nationale au besoin (Commission européenne, 2021b). Plusieurs pays de l'UE — Chypre, le Danemark, l'Italie, la Lettonie, Luxembourg, Malte, les Pays-Bas, la Pologne, le Portugal, la Slovaquie et la Suède — ont explicitement indiqué que leur politique et leur stratégie en matière d'IA seraient régulièrement examinées et mises à jour, généralement chaque année. De plus, la France a signalé son intention de mettre sur pied une plateforme nationale de vérification des algorithmes d'IA (Van Roy, 2019). La Commission européenne exige également que les fournisseurs de « systèmes d'IA à haut risque »²⁸ mettent au point des processus et des indicateurs de suivi de la conception, du développement et du test de ces systèmes, qui doivent être régulièrement vérifiés par un « organisme notifié », à titre de contrôle qualité (Commission européenne, 2021b).

Il sera également important que les gouvernements examinent et modifient périodiquement leurs politiques compte tenu de leurs implications pratiques et des progrès réalisés dans le domaine de l'IA. Ce principe sera également vrai pour les autres parties prenantes importantes, telles que les universités, les organismes de financement et d'industrie. Pour cela, les décideurs devront être conscients de la variabilité des échelles de temps des effets de leurs politiques, les conséquences des politiques concernant le financement de la recherche et l'innovation pouvant se matérialiser plus lentement que celles concernant les approvisionnements ou régissant l'utilisation de l'IA dans la recherche intra-muros. La participation multisectorielle aux examens peut également être bénéfique pour la diffusion de l'IA en sensibilisant aux défis et aux possibilités et en évitant que la réglementation semble être établie par un bassin restreint de parties prenantes (section 2.3).

28 Dans la réglementation de l'UE, les systèmes d'IA à haut risque sont les systèmes « qui présentent des risques importants pour la santé, la sécurité et les droits fondamentaux des personnes » (Commission européenne, 2021b).

Les décisions prises à l'extérieur du Canada créeront une pression pour harmoniser les approches canadiennes avec celles mises en œuvre à l'étranger

Malgré son leadership historique dans certains secteurs de la recherche sur l'IA, le Canada a du mal à élaborer son cadre réglementaire isolément de ses plus gros voisins et partenaires commerciaux. À cet égard, bien que la Directive sur la prise de décisions automatisée représentait une importante étape proactive vers la définition de normes et de pratiques à l'échelle de la fonction publique, les cadres réglementaires mis en place dans d'autres pays devraient influencer sur de nombreuses parties prenantes au Canada. Jusqu'ici, les voies empruntées pour réglementer l'IA dans d'autres pays phares de la R-D en IA sont diverses, et s'étendent de démarches menées par le marché et reposant sur l'autorégulation des développeurs à des approches dans lesquelles le gouvernement exerce un rôle plus prescriptif. L'autorégulation s'effectue selon des lignes directrices ou des normes qui pourraient être définies par l'industrie elle-même ou par des groupes consultatifs nationaux (Geist, 2021) (section 2.3). Plusieurs pays, comme le Canada, Israël, Royaume-Uni et les États-Unis, adoptent une approche guidée par le marché — jusqu'ici à divers degrés — alors que la Chine maintient un modèle centré sur le gouvernement (Geist, 2021).

Ce sont les États-Unis qui ont le plus résolument adopté la posture menée par le marché, agissant selon les principes du droit souple et de l'autorégulation décrits dans des principes directeurs et des codes de conduite pour chaque partie prenante (Castets-Renard, 2021). Il n'existe pas d'exemple de lois régissant les systèmes d'IA à l'échelle fédérale²⁹, celle-ci ayant tendance à s'appliquer à des secteurs particuliers que l'on retrouve seulement à l'échelle de l'État et de la municipalité, par exemple le transport et l'application de la loi, respectivement (Castets-Renard, 2021). Bien qu'il soit encadré par le gouvernement, le modèle chinois est fortement décentralisé dans la pratique en raison du nombre de paliers gouvernementaux concernés (Roberts *et al.*, 2021). Son cadre met davantage l'accent sur l'innovation, le développement économique et les avantages sociétaux, et moins sur les droits individuels ou sur la sécurité (Roberts *et al.*, 2021).

En 2021, l'UE a dévoilé sa proposition de réglementation de l'IA, qui se situe entre les modèles réglementaires américain et chinois (Geist, 2021). La réglementation se veut technologiquement neutre et s'applique aux entreprises publiques et privées, ainsi qu'aux systèmes de PDA et au contenu généré par l'IA (Commission européenne, 2021b). Elle est fondée sur les risques, tout comme le RGPD, dans lequel ces derniers sont définis en termes de sécurité, de droits humains, d'incertitude et de spécificité (Commission européenne, 2021a). Elle attribue un niveau de gravité à ces risques et

²⁹ Des initiatives fédérales ont été lancées aux États-Unis, telles que le National AI Initiative Office (Niczpeyr, 2021), mais la législation qui l'accompagne s'est jusqu'à présent concentrée sur l'orientation du développement de l'IA et non sur les instruments réglementaires (NAIIO, 2021).

impose ainsi des exigences proportionnelles aux développeurs et aux fournisseurs. Les stratégies d'atténuation des risques reposent sur la gouvernance des données et la documentation, et l'UE prévoit utiliser des bacs à sable réglementaires pour donner aux PME l'occasion de tester le nouveau cadre dans le contexte de systèmes d'IA à faible risque (Marcia et DeSouza, 2021). Les applications à haut risque concernant les systèmes disposant d'une forte autonomie exigeront des mécanismes de supervision rigoureux (p. ex. le régime de « responsabilité stricte » proposé, qui est examiné à la section 6.3). Floridi (2021) avertit cependant que la réglementation ne fait pas suffisamment la distinction entre le risque découlant d'une défaillance ou d'une erreur et le risque dû à l'intention de l'utilisateur.



« La diversité des priorités fédérales et provinciales ou territoriales pourrait aussi conduire à un environnement de réglementation de l'IA souple, mais en fin de compte fragmenté. »

Première à dévoiler une proposition de cadre, l'UE espère affirmer son leadership en matière de développement responsable de l'IA (Commission européenne, 2021a). Néanmoins, l'adoption d'une position réglementaire forte est risquée, étant donné l'absence d'harmonisation internationale. Par exemple, l'UE pourrait involontairement diriger le développement d'applications d'IA problématiques sur le plan éthique ailleurs, lesquelles pourraient être par la suite importées après avoir été rendues conformes dans leur pays d'origine (Floridi, 2021).

Cet environnement réglementaire international évolutif a une incidence sur l'activité canadienne d'IA en sciences et en génie, en raison de la divergence d'approches entre certains des plus gros partenaires du Canada en recherche comme en commerce (Geist, 2021). Cette situation pourrait pousser le Canada à harmoniser son approche avec celle d'autres pays afin de minimiser les incohérences. Face à ce problème, les gouvernements fédéral et provinciaux ou territoriaux pourraient tenter de copier la réglementation proposée par l'UE, en mettant l'accent sur l'importance des droits et de la prévention des préjudices sociétaux (Castets-Renard, 2021). La diversité des priorités fédérales et provinciales ou territoriales pourrait aussi conduire à un environnement de réglementation de l'IA souple, mais en fin de compte fragmenté. Dawson (2021) suggère que le Canada pourrait pourtant mener une initiative multilatérale visant à harmoniser les réglementations et les normes, mais tant que cette initiative n'aura pas été lancée, les chercheurs en IA canadiens devront être conscients du point auquel la divergence des normes internationales, surtout en ce qui concerne la gouvernance des données, influe sur leurs collaborations. De même, les PME canadiennes (qui peuvent se sentir poussées à accéder à certains marchés) devront atteindre un équilibre entre la conformité avec les réglementations internationales et l'incertitude nationale.

Conclusion

7.1 Réponse au mandat

7.2 Réflexions du comité d'experts

La concrétisation des promesses d'intégration de l'IA avec les sciences et le génie permettra d'examiner de nouvelles questions et accélérera l'innovation dans de nombreux autres secteurs technologiques. Ce rapport étudie le vaste éventail de questions légales et réglementaires, éthiques, sociales et politiques reliées à la conception et au déploiement de l'IA en sciences et en génie.

7.1 Réponse au mandat

Les progrès continus des algorithmes et la grande disponibilité des ressources de calcul et des données scientifiques offrent les conditions idéales pour l'application de l'IA à la conception et à la découverte en sciences et en génie. Cependant, malgré les percées accrues qui démontrent les promesses que recèle cette technologie, plusieurs défis réels et imminents doivent tout d'abord être surmontés. Afin de mieux comprendre les possibilités, les défis et les conséquences, le CNRC, appuyé par le CIFAR, les IRSC, le CNRSNG et le CRSH, a demandé au CAC de constituer un comité multidisciplinaire et multisectoriel d'experts pour répondre à la question suivante :



Quels sont les défis légaux/réglementaires, éthiques, sociaux et politiques du déploiement des technologies d'IA pour permettre la conception de la recherche et la découverte en sciences et en génie au Canada?

Dans cette section, le comité d'expert répond à cette question et souligne ses principales conclusions.

L'utilisation accrue de l'IA transformera les pratiques épistémiques, éthiques et institutionnelles en sciences et en génie

Surmonter les défis relatifs à l'exactitude, à l'explicabilité et à la reproductibilité des résultats générés par le système d'IA nécessitera une meilleure transparence de la part des chercheurs et l'établissement de nouvelles normes et pratiques de gestion des données pour protéger l'intégrité scientifique. De plus, même si les résultats eux-mêmes sont fiables, la complexité et l'opacité de certains types de systèmes mènent à s'interroger si les utilisateurs seront en mesure de comprendre ou d'expliquer les conclusions ou les concepts nouveaux qui pourraient être obtenus à l'aide de ces systèmes.

Les questions éthiques à propos de l'utilisation de l'IA en sciences et en génie surgissent à tous les stades du processus, que ce soit au recueil et au prétraitement des données, à la conception et au déploiement des modèles d'IA entraînés avec ces données, à la diffusion et à la publication des résultats ou encore au stockage, à la tenue à jour et à l'accès à long terme aux données, modèles et résultats. L'importance des données, et en particulier des mégadonnées, crée des défis pour de nombreux aspects de l'éthique de la recherche traditionnelle, en particulier lorsque celle-ci fait appel à des participants humains et à des groupes historiquement marginalisés. Pour régler ces problèmes, les acteurs



« Même si les résultats eux-mêmes sont fiables, la complexité et l'opacité de certains types de systèmes mènent à s'interroger si les utilisateurs seront en mesure de comprendre ou d'expliquer les conclusions ou les concepts nouveaux qui pourraient être obtenus à l'aide de ces systèmes. »

institutionnels (y compris les organismes de financement, les universités et les entreprises de R-D) peuvent avoir à modifier leurs cadres de conduite éthique de la recherche afin de tenir compte des implications de l'IA. Par exemple, l'utilisation de l'IA en sciences et en génie peut avoir une incidence sur des groupes sociaux plus que sur les individus, ce qui exige de moins se concentrer sur les préjudices individuels et davantage sur les préjudices sociaux. Cependant, les comités d'examen éthique de la recherche traditionnels sont mal outillés pour évaluer ces répercussions.

L'intensification de l'utilisation de l'IA dans les domaines des sciences et du génie pourrait également changer leur dynamique sociale. Par exemple, l'augmentation de l'emploi de l'IA dans la recherche scientifique pourrait nécessiter la révision des pratiques sociales sous-tendant la diffusion des conclusions scientifiques (publications, conférences), l'établissement de leur validité (examen par les pairs, réplication) et la mention de leur provenance (crédit, citations). L'accroissement du recours à l'IA transformera la nature de la recherche scientifique et donc, exigera des scientifiques humains et des ingénieurs qu'ils travaillent différemment. Il est possible que permettre aux machines d'orienter la recherche future aide à surmonter certaines limites de la prise de décision humaine, mais au détriment du contrôle.

Les approches mettant l'accent sur l'accès équitable, la diversité et l'inclusion aideront à surmonter les défis sociaux et éthiques liés à l'utilisation de l'IA dans le système de recherche canadien

Tant le domaine de la recherche sur l'IA lui-même que l'accès à la technologie présentent d'importantes iniquités. Le manque actuel de diversité de genre et de race dans le domaine est bien documenté, et il y a une forte inégalité dans la répartition des ressources, de l'infrastructure et des compétences relatives à l'IA en recherche scientifique. Garantir l'accès équitable à l'IA est bénéfique à de nombreux égards et constitue un élément crucial dont les gouvernements et les établissements doivent s'occuper. D'une part, améliorer l'accès aux ressources nécessaires pour développer ou appliquer des systèmes d'IA en sciences et en génie facilitera la diffusion de la technologie dans les domaines, avec la promesse d'ouvrir sur un vaste éventail d'applications. D'autre part, l'accès est également un élément essentiel pour la répartition des avantages offerts par les systèmes d'IA déjà développés. Dans les deux cas, de solides encouragements à l'inclusion et à la transparence amélioreront l'équité dans la recherche sur l'IA et dans l'accès à cette technologie au Canada. Toutefois, l'exacerbation des inégalités dans l'accès aux ressources d'IA peut être difficile à éviter, étant donné le degré d'influence que les adopteurs précoces peuvent exercer sur son développement rapide. Les lacunes de l'environnement canadien de la recherche et de l'innovation risquent donc de s'agrandir si on ne prend pas soin de promouvoir l'accès aux outils, aux ressources et à l'emploi dans ce secteur émergent.

Les organismes de financement de la recherche peuvent devoir adapter certaines de leurs pratiques d'évaluation parce que l'allocation de fonds de recherche publics dictera les types d'applications d'IA en sciences et en génie qui seront priorisées et les normes de conduite auxquelles on s'attendra des bénéficiaires du financement. Le système de financement de la recherche peut également devoir intégrer des outils d'IA dans ses processus d'évaluation de la recherche et de ses répercussions. Il devra également mettre sur pied des programmes et des politiques pour entretenir de nouveaux partenariats interdisciplinaires et promouvoir la création et le partage de données de grande qualité. Par conséquent, il sera particulièrement important pour les organismes de financement de faire le suivi des questions entourant le biais et la discrimination lors de l'adoption de l'IA dans leurs processus de gestion du financement. La crédibilité de ces initiatives reposera sur l'établissement de pratiques semblables pour la R-D afin de garantir l'exactitude, d'interpréter l'incertitude et de promouvoir la confiance.

Comme elle s'effectue au sein d'un environnement légal et réglementaire incertain, l'utilisation de l'IA en sciences et en génie soulève des questions sur ses avantages et ses préjudices sociaux

L'intensification de l'utilisation des outils d'IA en sciences et en génie aura probablement d'importantes conséquences sur la société dans son ensemble. Bien que les découvertes effectuées au moyen de systèmes d'IA pourraient aider à lutter contre la crise climatique, les graves impacts environnementaux du développement et de l'utilisation de cette technologie demeurent une crainte non dissipée. De plus, l'automatisation des sciences et du génie résultant de l'intégration de l'IA pourrait être vue comme une menace pour le marché du travail de la recherche au Canada parce qu'au moins certains postes dans ces secteurs pourraient être éliminés par cette technologie. Les systèmes d'IA ont également servi d'amplificateur des iniquités et problèmes sociaux existants dans de nombreux secteurs où ils ont



« L'automatisation des sciences et du génie résultant de l'intégration de l'IA pourrait être vue comme une menace pour le marché du travail de la recherche au Canada parce qu'au moins certains postes dans ces secteurs pourraient être éliminés par cette technologie. »

été déployés jusqu'ici. La manifestation répétée de résultats discriminatoires contre des groupes historiquement marginalisés est une faiblesse bien connue et dangereuse de ces systèmes et pourrait être perpétuée par leur utilisation en sciences et en génie. Par exemple, les communautés autochtones peuvent devoir participer à des collaborations scientifiques utilisant des systèmes d'IA qui ont été conçus sans leur rapport, avec le risque que les pratiques de conduite et de données de recherche culturellement appropriées ne soient pas suivies.

Certains de ces risques sont dus aux écarts entre les principes de développement responsable de l'IA et leur opérationnalisation, ainsi que par l'insuffisance globale de solides mesures réglementaires. Tant au Canada qu'à l'étranger, l'environnement légal et

réglementaire a actuellement du mal à suivre le rythme des progrès technologiques et le désir de faciliter l'innovation se heurte au devoir de protéger la société contre les préjudices. Ce conflit se manifeste de nombreuses façons, que ce soit par les compromis réalisés entre la protection et le partage des données scientifiques ou par les décisions d'attribution de la responsabilité concernant les systèmes autonomes. Cette situation reflète également les difficultés à atteindre un consensus sur le statut juridique des découvertes et des concepts produits par les systèmes d'IA, ce qui a des conséquences sur la commercialisation et l'accès. Dans les domaines où la réglementation fait actuellement défaut, les contrats et les licences peuvent aider à délimiter les droits et les responsabilités des innovateurs. Enfin, les efforts pour harmoniser et moderniser les cadres légaux et

réglementaires à l'échelon national et international se heurtent parfois à des obstacles dus à la diversité des valeurs culturelles et politiques et à la dépendance de sentier, ce qui conduit à un paysage hétérogène qui influence le moment et l'endroit où les applications sont développées.

Les politiques et stratégies d'IA nationales qui transcendent les domaines de politique peuvent stimuler les développements en sciences et en génie

Un nombre croissant de pays, dont le Canada, ont élaboré des stratégies nationales concernant l'IA. La stratégie canadienne s'est généralement concentrée sur la recherche et sur le renforcement des capacités, mais a récemment commencé à cibler plusieurs secteurs et tient compte plus explicitement des impacts sociétaux. L'élargissement de la portée des stratégies d'IA peut mener à de nouvelles relations au sein du réseau de R-D et à établir des liens cruciaux entre la politique d'IA et des domaines extérieurs à son élaboration. Par exemple, les politiques promouvant l'accès aux données et aux données ouvertes ne sont pas strictement des politiques d'IA, elles sont des éléments facilitateurs importants. Les investissements



« L'environnement légal et réglementaire a actuellement du mal à suivre le rythme des progrès technologiques et le désir de faciliter l'innovation se heurte au devoir de protéger la société contre les préjudices. »

dans l'infrastructure de données et les initiatives de politique sur les données se multiplient de façon correspondante, soit à titre de composant explicite des récentes stratégies d'IA internationales soit en complément à celles-ci.

Les décisions et les avancées dans de nombreuses dimensions politiques auront une incidence sur la vitesse et sur la trajectoire de la croissance de l'utilisation de l'IA en sciences et en génie. En fait, certaines recherches scientifiques utilisant l'IA coïncident déjà avec les objectifs et les initiatives politiques existants : les efforts pour lutter contre les changements climatiques par l'innovation, par exemple, encouragent le financement de la recherche en sciences des matériaux au moyen

de l'IA. D'autres tendances politiques, telles que la transformation des méthodes d'évaluation de la recherche et des chercheurs, sont annexes à l'IA, mais elles influenceront aussi son développement. Pour faire progresser l'IA en sciences et en génie, il sera utile de déterminer et de gérer proactivement les interconnexions entre les secteurs politiques pour tenir compte de la grande variété de parties prenantes au-delà de la communauté de l'IA.

7.2 Réflexions du comité d'experts

La recherche et les commentaires sur le développement et l'utilisation responsables de l'IA se concentrent généralement sur les applications susceptibles de nuire aux droits humains ou à la protection de la vie privée. Par comparaison, on s'est moins penché sur son développement et son utilisation dans la recherche en sciences et en génie, bien que les préoccupations qui émergent dans ces disciplines ressemblent à celles déjà prises en compte pour la société dans son ensemble. De plus, bien que l'innovation en sciences et en génie soit parfois perçue comme intrinsèquement bonne, les progrès dans la technologie d'IA pour ces

disciplines conduiront aussi à des avancées de l'IA dans d'autres secteurs plus controversés.

L'établissement de bonnes pratiques et la lutte contre l'amplification des problèmes sociaux et éthiques par les systèmes d'IA utilisés en sciences et en génie peuvent fournir des leçons sur la façon d'éviter les effets indésirables dans d'autres pans de la société dans lesquels l'IA est déployée.

La nature intersectorielle de l'IA souligne le besoin d'acquérir une expertise et de développer des rôles pour mieux intégrer les connaissances et les compétences dans plusieurs disciplines traditionnelles. L'expertise en gestion des données, ou en examen des outils d'IA, alliée à des politiques encourageant la transparence, aidera à résoudre plusieurs défis épistémiques et éthiques au stade de la conceptualisation et du développement des outils d'IA pour les sciences et le génie. De même,

le système de financement de la recherche peut mieux soutenir la R-D en IA au Canada en s'adaptant et en apprenant à évaluer la recherche et les chercheurs selon différents prismes. Par exemple, le concept d'*excellence de la recherche* pourrait devoir être revu si les systèmes d'IA assument des rôles plus étoffés dans l'orientation de la recherche. La formation d'universitaires et d'ingénieurs capables de comprendre les dilemmes éthiques et sociaux pourrait prévenir les conséquences non éthiques ou involontaires et, donc, renforcer la confiance du public dans les découvertes effectuées grâce à l'IA ou dans les objets, matériaux et processus conçus par l'IA. Dans un même temps, le déploiement continuera à être compliqué par les incertitudes persistantes concernant plusieurs dimensions légales et réglementaires non résolues. Les décideurs bénéficieraient d'une plus vaste consultation entre les secteurs gouvernementaux et professionnels dans leurs tentatives pour faire des choix en matière de gouvernance des données,



« La formation d'universitaires et d'ingénieurs capables de comprendre les dilemmes éthiques et sociaux pourrait prévenir les conséquences non éthiques ou involontaires et, donc, renforcer la confiance du public dans les découvertes effectuées grâce à l'IA. »

de niveau acceptable de risque sociétal et de façon de gérer les droits de propriété intellectuelle. Bien que des réformes dans ce dernier domaine soient en cours, des examens et un renouvellement réguliers des politiques actuelles relatives à l'IA dans tous les domaines aideront à éviter les résultats négatifs (ou à apprendre de ceux-ci) et à tirer profit des nouvelles possibilités.

Ce rapport cerne le grand nombre de parties prenantes dont les actions et les décisions détermineront comment ces défis seront surmontés et qui façonnera la façon dont des champs et des secteurs disparates pourraient intégrer l'IA dans leurs pratiques. Cette situation fluide offre l'occasion de se tourner vers l'avenir et définir les domaines le long du cycle de vie du développement et du déploiement de l'IA dans lesquels les décideurs pourraient intervenir avant que les obstacles prévus ne se matérialisent graduellement. En premier lieu, les organismes de financement de la recherche, les établissements universitaires et les organismes publics chargés de soutenir l'innovation peuvent entreprendre des démarches



« Cette situation fluide offre l'occasion de se tourner vers l'avenir et définir les domaines le long du cycle de vie du développement et du déploiement de l'IA dans lesquels les décideurs pourraient intervenir avant que les obstacles prévus ne se matérialisent graduellement. »

proactives pour atténuer les barrières à la collaboration interdisciplinaire et intersectorielle entre les chercheurs et les spécialistes de l'IA en sciences et en génie et les personnes qui travaillent dans d'autres disciplines. L'ajustement des pratiques de financement, de la formation et de l'apprentissage et de l'accès aux ressources d'IA contribuera à la capacité du Canada à tirer profit des forces dans la recherche en IA acquises au cours des précédentes décennies pour utiliser cette technologie dans la conception et la découverte en sciences et en génie. L'incapacité à adapter les pratiques existantes dans ces domaines, ainsi que dans d'autres, compliquera les efforts d'élargissement de l'écosystème d'IA canadien au-delà des actuels centres majeurs dans lesquels la R-D est la plus active, avec la possibilité de manquer des occasions de croissance horizontale

à l'extérieur de la recherche fondamentale en IA. Après tout, les actuels réseaux qui composent le milieu de la R-D au Canada n'évolueront pas nécessairement de façon à assurer l'usage le plus efficace de l'IA en sciences et en génie sans de nouvelles relations, voire des relations inattendues. Le degré auquel la société canadienne pourra bénéficier des décennies d'investissements publics dans l'IA reposera sur ces relations et sur l'adoption qui en résultera des outils d'IA dans l'ensemble de la communauté de l'innovation.

Il est toutefois important de se rendre compte que pour profiter des avantages d'employer l'IA dans la recherche en sciences et en génie, il faudra faire en sorte que ces systèmes soient utilisés avec soin et de façon responsable et judicieuse.



« Le degré auquel la société canadienne pourra bénéficier des décennies d'investissements publics dans l'IA reposera sur ces relations et sur l'adoption qui en résultera des outils d'IA dans l'ensemble de la communauté de l'innovation. »

Si l'IA offre la possibilité de transcender les limites des capacités cognitives humaines en produisant des découvertes scientifiques et des concepts de génie novateurs, elle a aussi le potentiel de perpétuer des biais humains, et d'en créer de tout nouveaux. Si elle n'est pas utilisée de manière responsable, elle pourrait exacerber les inégalités dans les systèmes de recherche et d'innovation, ainsi que dans l'ensemble de la société. Pour éviter de tels effets négatifs, il faudra tenir compte des facteurs sociaux et éthiques, non seulement à l'étape du déploiement de l'IA, mais aussi aux toutes premières étapes du développement. De plus, l'utilisation responsable de l'IA en

sciences et en génie exigera également de la transparence et des mécanismes de supervision afin que les résultats générés par l'IA soient exacts, reproductibles et explicables, de façon à améliorer le savoir scientifique.

Références

- Acceleration Consortium, 2021. Mission. Adresse : <https://acceleration.utoronto.ca/#mission> (consulté en juin 2021).
- Adadi, A. et M. Berrada, 2018. « Peeking inside the black-box: A survey on explainable artificial intelligence (XAI) », *IEEE Access*, vol. 6, p. 52138–52160.
- Amit, R. et R. Meir, 2019. « Lifelong learning and inductive bias », *Current Opinion in Behavioral Sciences*, vol. 29, p. 51–54.
- Amnesty International et Access Now, 2018. *Déclaration de Toronto Protéger le droit à l'égalité et à la non-discrimination dans les systèmes reposant sur l'apprentissage automatique*, Toronto, ON, Amnesty International and Access Now.
- Ananthaswamy, A., 2021a. « Latest Neural Nets Solves Worlds Hardest Equations Faster than Ever Before ». *Quanta Magazine* (19 avril).
- Ananthaswamy, A., 2021b. « AI Designs Quantum Physics Experiments Beyond What Any Human Has Conceived ». *Scientific American* (2 juillet).
- Anthony, L. F. W., B. Kanding, et R. Selvan, 2020. « Carbontracker – Tracking and Predicting the Carbon Footprint of Training Deep Learning Models [pre-print] », *arXiv*, n°2007.03051v1.
- Anyaha, R., 2017. The History of Artificial Intelligence. Adresse : <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/> (consulté en mai 2021).
- ARDC – Australian Research Data Commons, 2021. Research Data Australia. Adresse : <https://researchdata.edu.au/page/about> (consulté en juillet 2021).
- Asay, C. D., 2020. « Independent Creation in a World of AI », *Florida International University Law Review*, vol. 14, n°2.
- Assemblée Nationale (Assemblée Nationale du Québec), 2020. *Projet de loi no 64 — Loi Modernisant des Dispositions Législatives en Matière de Protection des Renseignements Personnels*, Québec, QC, Gouvernement du Québec.
- Baker, D., D. Bourne-Tyson, L. Gerlitz, S. Haigh, S. Khair, M. Leggott, ... M. Whitehead, 2019. *Research Data Management in Canada: A Backgrounder*, Canadian Association for Research Libraries (CARL), Consortia Advancing Standards in Research Administration Information (CASRAI), Leadership Council for Digital Research Infrastructure (LCDRI), Research Data Canada (RDC).
- Bates, J., 2021. Expanding the Geography of Innovation: NSF AI Institutes 2021. Adresse : <https://beta.nsf.gov/science-matters/expanding-geography-innovation-nsf-ai-research-institutes-2021> (consulté en août 2021).
- Bawa, K. et M. Tawfik, 2020. « How to Keep IP Canadian: Learn from the Success of Public-private Partnerships During the Pandemic ». *The Globe and Mail* (30 décembre).

- BCOHRC – British Columbia's Office of the Human Rights Commissioner, 2020. *Disaggregated Demographic Data Collection in British Columbia: The Grandmother Perspective*, Vancouver (BC), BCOHRC.
- Bejani, M. M. et M. Ghaate, 2021. « A systematic review on overfitting control in shallow and deep neural networks », *Artificial Intelligence Review*.
- Benhamou, Y. et J. Ferland, 2021. « Artificial Intelligence and Damages: Assessing Liability and Calculating Damages », dans D'Agostino, G., A. Gaon et C. Piovesan (éd.), *Leading Legal Disruption: Artificial Intelligence and a Tool Kit for Lawyers and the Law*, Toronto, ON, Thomson Reuters Canada.
- Berendt, B. et S. Preibusch, 2017. « Toward accountable discrimination-aware data mining: The importance of keeping the human in the loop—and under the looking glass », *Big Data*, vol. 5, n°2, p. 135-152.
- Berman, F., 2019. *The Research Data Alliance — The First Five Years*, Research Data Alliance.
- Bernstein, M. S., M. Levi, D. Magnus, B. Rajala, D. Satz, et C. Waeiss, 2021. « ESR: Ethics and Society Review of artificial intelligence research », *arXiv*, n°2106.11521v2.
- Bezuidenhout, L. et E. Ratti, 2020. « What does it mean to embed ethics? », *AI & Society*, n°36, p. 939-953.
- Birhane, A., 2021. « Algorithmic injustice: A relational ethics approach », *Patterns*, vol. 2, n°2, p. 100205.
- Boeckhout, M., G. A. Zielhuis, et A. L. Bredenoord, 2018. « The FAIR guiding principles for data stewardship: Fair enough? », *European Journal of Human Genetics*, vol. 26, n°7, p. 931-936.
- Bogue, R., 2012. « Robots in the laboratory: A review of applications », *Industrial Robot*, vol. 39, n°2, p. 113-119.
- Bond, T., 2021. News Flash: UK Court of Appeal Rejects DABUS Patent Application. Adresse : <https://www.twobirds.com/en/news/articles/2021/uk/uk-court-of-appeal-rejects-dabus-patent-application> (consulté en octobre 2021).
- Borrows, J., 2010. *Canada's Indigenous Constitution*. Toronto, ON, University of Toronto Press.
- Borrows, J., 2012. « (Ab)Originalism and Canada's constitution », *The Supreme Court Law Review: Osgoode's Annual Constitutional Cases Conference*, vol. 58, p. 351-398.
- Boschert, S., C. Heinrich, et R. Rosen, 2018. *Next Generation Digital Twin*, communication présentée dans le cadre du Twelfth International Symposium on Tools and Methods of Competitive Engineering, Las Palmas de Gran Canaria, Espagne.
- Boulanin, V., K. Brockmann, et L. Richards, 2020. *Responsible Artificial Intelligence Research and Innovation for International Peace and Security*, Solna, Suède, Stockholm International Peace Research Institute.
- Boutilier, A., 2021. « RCMP Broke Privacy Laws in Using Controversial Clearview AI Facial Recognition Tools, Watchdog Says ». *Toronto Star* (10 juin).

- boyd, d. et K. Crawford, 2012. « Critical questions for big data », *Information, Communication & Society*, vol. 15, n°5, p. 662-679.
- Brandusescu, A., 2021. *Artificial Intelligence Policy and Funding in Canada: Public Investments, Private Interests*, Montréal, QC, Centre for Interdisciplinary Research on Montreal.
- Brandusescu, A., A. Cutean, P. Dawson, R. Davidson, M. Matthews, et K. O'Neill, 2021. *Maximizing Strength and Spearheading Opportunity: An Industrial Strategy for Canadian AI*, Ottawa, ON, Information and Communications Technology Council.
- Brenner, S., 1995. « Loose ends », *Current Biology*, vol. 5, p. 568.
- Brodt, O., M. Khavkin, L. Rokach, A. Shabtai, et Y. Elovici, 2021. « Artificial Intelligence and (the Lack of) Security », dans D'Agostino, G., A. Gaon et C. Piovesan (éd.), *Leading Legal Disruption: Artificial Intelligence and a Tool Kit for Lawyers and the Law*, Toronto, ON, Thomson Reuters Canada.
- Brynjolfsson, E., T. Mitchell, et D. Rock, 2018. « What can machines learn and what does it mean for occupations and the economy? », *AEA Papers and Proceedings*, vol. 108, p. 43-47.
- CAC – Conseil des académies canadiennes, 2012. *Renforcer la capacité de recherche du Canada : La dimension de genre*, Ottawa, ON, Le comité d'experts sur les femmes dans la recherche universitaire, CAC.
- CAC – Conseil des académies canadiennes, 2013. *Paradoxe dissipé*, Ottawa, ON, CAC.
- CAC – Conseil des académies canadiennes, 2018. *Rivaliser dans une économie mondiale axée sur l'innovation : L'état de la R-D au Canada*, Ottawa, ON, Le comité d'experts sur l'état de la science et de la technologie et de la recherche-développement industrielle au Canada, CAC.
- CAC – Conseil des académies canadiennes, 2021a. *Dynamiser la découverte*, Ottawa, ON, Le comité d'experts sur les pratiques internationales pour financer la recherche en sciences naturelles et en génie, CAC.
- CAC – Conseil des académies canadiennes, 2021b. *En attente de connexion*, Ottawa, ON, Le comité d'experts sur les réseaux à haut débit pour les collectivités rurales et éloignées, CAC.
- CAC – Conseil des académies canadiennes, 2010. *Honnêteté, responsabilité et confiance : Promouvoir l'intégrité en recherche au Canada*, Ottawa, ON, Le comité d'experts sur l'intégrité en recherche, CAC.
- Carroll, M. W., 2019. « Copyright and the progress of science: Why text and data mining is lawful », *UC Davis Law Review*, vol. 53, p. 893-963.
- Carroll, S. R., E. Herczog, M. Hudson, K. Russell, et S. Stall, 2021. « Operationalizing the CARE and FAIR Principles for Indigenous data futures », *Scientific Data*, vol. 8, n°1, p. 108.
- Carroll, S. R., I. Garba, O. L. Figueroa-Rodríguez, J. Holbrook, R. Lovett, S. Materechera, ... M. Hudson, 2020. « The CARE principles for Indigenous data governance », *Data Science Journal*, vol. 19, n°43, p. 1-12.

- Carter, L., D. Liu, et C. Cantrell, 2020. « Exploring the intersection of the digital divide and artificial intelligence: A hermeneutic literature review », *AIS Transactions on Human-Computer Interaction*, vol. 12, n°4, p. 253-275.
- Carter, R. E., Z. I. Attia, F. Lopez-Jimenez, et P. A. Friedman, 2019. « Pragmatic considerations for fostering reproducible research in artificial intelligence », *Nature Partner Journals Digital Medicine*, vol. 2, p. 42.
- Castelvecchi, D., 2021. « AI maths whiz creates tough new problems for humans to solve », *Nature*, vol. 590, p. 196.
- Castets-Renard, C., 2021. « AI and the Law in the European Union and the United States », dans Martin-Bariteau, F. et T. Scassa (réd.), *Artificial Intelligence and the Law in Canada*, Toronto, ON, LexisNexis.
- CCN – Conseil canadien des normes, 2020. Le Collectif canadien de normalisation en matière de gouvernance des données. Adresse : <https://www.scc.ca/fr/programmes-phares/gouvernance-des-donnees> (consulté en mai 2021).
- CDO – Commission du drot de l'Ontario, 2021. *Réglementer l'intelligence artificielle – Enjeux et choix essentiels*, Toronto, ON, Osgoode Law School, York University.
- Cecaroni, L., J. Bibby, E. Roger, P. Flemons, K. Michael, L. Fagan, et J. L. Oliver, 2019. « Opportunities and risks for citizen science in the age of artificial intelligence », *Citizen Science: Theory and Practice*, vol. 4, n°1, p. 29.
- CEST – Commission de l'Éthique en Science et en Technologie, 2021. *Mémoire de la Commission de l'Éthique en Science et en Technologie — Consultations particulières sur le projet de loi no 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels*, Québec, QC, Gouvernement du Québec.
- CGIPN – Le Centre de gouvernance de l'information des Premières Nations, s.d. À propos de nous. Adresse : <https://fnigc.ca/fr/a-propos-de-nous/> (consulté en août 2021).
- Chagal-Feferkorn, K. A., 2019. « Am I an algorithm or a product? When products liability should apply to algorithmic decision-makers », *Stanford Law & Policy Review*, vol. 30, p. 61-113.
- Chang, T. Y.-C., 2021. DABUS the AI is Denied Inventorship in Taiwan. Adresse : <https://www.mondaq.com/patent/1108642/dabus-the-ai-is-denied-inventorship-in-taiwan> (consulté en octobre 2021).
- Charlin, L. et R. S. Zemel, 2013. *The Toronto Paper Matching System: An Automated Paper-Reviewer Assignment System*, communication présentée dans le cadre du 30th International Conference on Machine Learning, Atlanta, GA.
- Chawla, D. S., 2020. « Do You Have A Conflict Of Interest? This Robotic Assistant May Find It First ». *The New York Times* (novembre 23).
- Chawla, D. S., 2021. « Frosty Reception For Algorithm That Predicts Research Papers' Impact ». *Nature* (21 mai).

- Checco, A., L. Bracciale, P. Loreti, S. Pinfield, et G. Bianchi, 2021. « AI-assisted peer review », *Humanities and Social Sciences Communications*, vol. 8, n°1, p. 25.
- Chen, M., J. Twork, H. Jun, Q. Yuan, H. Ponde de Oliveira Pinto, J. Kaplan, ... W. Zaremba, 2021. « valuating large language models trained on code », *arXiv*, n°2107.03374v2.
- Chowdhury, N., Z. Hakim, N. A. Taylor, T. Remennik, S. Rogers, E. Strome, et J. Wallis, 2020. *Pan-Canadian AI Strategy Impact Assessment Report*, Toronto, ON, Accenture & CIFAR.
- CIFAR – l’Institut canadien de recherches avancées, 2021. *IA et soins de santé : La fusion du droit et de la science*, Toronto, ON, CIFAR.
- CIPVP – La commissaire à l’information et à la protection de la vie privée de l’Ontario, 2016. *De-identification Guidelines for Structured Data*, Toronto, ON, CIPVP.
- Citron, D. K. et F. A. Pasquale, 2014. « The Scored Society: Due Process for Automated Predictions », *Washington Law Review*, vol. 89, 1-33.
- Ciuriak, D. et M. Ptashkina, 2021. *Quantifying Trade Secret Theft — Policy Implications*, Waterloo, ON, Centre for International Governance Innovation.
- CleverHans, 2021. CleverHans. Adresse : <https://github.com/cleverhans-lab/cleverhans> (consulté en août 2021).
- CNRC – Conseil national de recherches Canada, 2021. Programme Défi « L’intelligence artificielle au service de la conception ». Adresse : <https://nrc.canada.ca/fr/recherche-developpement/recherche-collaboration/programmes/programme-defi-lintelligence-artificielle-au-service-conception> (consulté en août 2021).
- cOAlition S, 2021. Principles and Implementation. Adresse : <https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/> (consulté en juillet 2021).
- CodeCarbon, 2021. CodeCarbon.io. Adresse : <https://codecarbon.io/> (consulté en août 2021).
- Cofone, I., A. Qarri, et J. Wiener, 2021. « Updating Canada’s Privacy Act for Artificial Intelligence ». *Policy Magazine* (4 février).
- Comité d’experts en matière de propriété intellectuelle, 2020. *Rapport : La propriété intellectuelle au sein du milieu de l’innovation de l’Ontario*, Toronto, ON, gouvernement de l’Ontario.
- Commission européenne, 2018. *Résumé de l’évaluation de la directive 96/9/CE concernant la protection des bases de données*, Bruxelles, Belgique, Commission Européenne.
- Commission européenne, 2021a. *Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l’intelligence artificielle (législation sur l’intelligence artificielle) et modifiant certains actes législatifs de l’union*, Bruxelles, Belgique, Commission Européenne.

- Commission européenne, 2021b. *Annexes à la Proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'union*, Bruxelles, Belgique, Commission Européenne.
- Cowls, J., A. Tsamadou, L. Floridi, et M. Taddeo, 2021. *The AI Gambit – Leveraging Artificial Intelligence to Combat Climate Change: Opportunities, Challenges, and Recommendations*, Berlin, Allemagne, Vodafone Institute for Society and Communications.
- CPVP – Commissariat à la protection de la vie privée du Canada, 2021a. Aperçu des lois sur la protection des renseignements personnels au Canada. Adresse : https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/lois-sur-la-protection-des-renseignements-personnels-au-canada/02_05_d_15/ (consulté en mai 2021).
- CPVP – Commissariat à la protection de la vie privée du Canada, 2021b. *Police Use of Facial Recognition Technology in Canada and the Way Forward: Special Report to Parliament on the OPC's Investigation into the RCMP's Use of Clearview AI and Draft Joint Guidance for Law Enforcement Agencies Considering the Use of Facial Recognition Technology*, Ottawa, ON, CPVP.
- CPVP – Commissariat à la protection de la vie privée du Canada, 2021c. *Enquête conjointe sur Clearview AI, Inc. par le Commissariat à la protection de la vie privée du Canada, la Commission d'accès à l'information du Québec, le Commissariat à l'information et à la protection de la vie privée de la Colombie-Britannique et le Commissariat à l'information et à la protection de la vie privée de l'Alberta*, Ottawa, ON, CPVP.
- Craig, C. et I. R. Kerr, 2021. « The death of the AI author », *Ottawa Law Review*, vol. 52, n°1, p. 31–86.
- Craig, C. J., 2021. « AI and Copyright », dans Martin-Bariteau, F. et T. Scassa (réd.), *Artificial Intelligence and the Law in Canada*, New York, NY, LexisNexis.
- Crawford, K., R. Dobbe, T. Dryer, G. Fried, B. Green, E. Kaziunas, ... M. Whittaker, 2019. *AI Now 2019 Report*, New York, NY, AI Now Institute.
- CRSH – Conseil de recherches en sciences humaines, 2021. Fonds Nouvelles frontières en recherche. Adresse : <https://www.sshrc-crsh.gc.ca/funding-financement/nfrf-fnfr/index-eng.aspx> (consulté en octobre 2021).
- CRSNG – Conseil de recherches en sciences naturelles et en génie du Canada, 2014. Foire aux questions sur la version préliminaire de la Politique de libre accès des trois organismes : À propos de la Politique. Adresse : https://www.nserc-crsng.gc.ca/NSERC-CRSNG/policies-politiques/OpenAccessFAQ-LibreAccesFAQ_fra.asp (consulté en septembre 2021).
- CSTI – Council for Science, Technology and Innovation, 2020. *Moonshot Goal 3*, Tokyo, Japon, Cabinet Office of Japan.
- Cussins Newman, J., 2019. *Towards AI Security: Global Aspirations for a more Resilient Future*, Berkeley, CA, Center for Long-Term Cybersecurity.

- Dastin, J., 2018. « Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women » *Reuters* (10 octobre).
- Davies, T., S. B. Walker, M. Rubinstein, et F. Perini (éd.), 2019. *The State of Open Data*. Cape Town, Afrique du Sud, African Minds et le Centre de recherches pour le développement international (CRDI).
- Dawson, P., 2021. « 'Ottawa Effect' Could Unlock Benefits of Global Co-operation on AI ». *The Globe and Mail* (23 mai).
- de Beer, J., 2020. *An International IP and Digital Trade Strategy for Canada*, Waterloo, ON, Centre for International Governance Innovation.
- De Luna, P., J. Wei, Y. Bengio, A. Aspuru-Guzik, et E. Sargent, 2017. « Use machine learning to find energy materials », *Nature*, vol. 552, n°7683, p. 23–27.
- Déclaration de Montréal IA responsable, 2021. Signataires. Adresse : <https://www.declarationmontreal-iaresponsable.com/signataires> (consulté en novembre 2021).
- Deep Genomics, 2021. AI-Powered Discovery. Adresse : <https://www.deepgenomics.com/> (consulté en septembre 2021).
- Deloitte Canada, 2019. *Canada's AI Imperative — Overcoming Risks, Building Trust*, Toronto, ON, Deloitte LLP.
- Deltorn, J.-M., 2017. « Deep creations: Intellectual property and the automata », *Frontiers in Digital Humanities*, vol. 4, n°3.
- DFDR – Dépôts de données de recherche au Canada, 2021. À propos. Adresse : https://www.fdr-dfdr.ca/docs/fr/%C3%AO__propos/ (consulté en juillet 2021).
- Dhar, P., 2020. « The carbon impact of artificial intelligence », *Nature Machine Intelligence*, vol. 2, p. 423–425.
- Dick, S., 2019. « Artificial intelligence », *Harvard Data Science Review*, vol., n°1.1.
- Dietvorst, B. J., J. P. Simmons, et C. Massey, 2015. « Algorithm aversion: People erroneously avoid algorithms after seeing them err », *Journal of Experimental Psychology*, vol. 144, n°1.
- Dignum, V., 2020a. « AI is multidisciplinary », *AI Matters*, vol. 5, n°4, p. 18–21.
- Dignum, V., 2020b. « Responsibility and Artificial Intelligence », dans Dubber, M. D., F. Pasquale et S. Das (éd.), *The Oxford Handbook of Ethics of AI*, Oxford, Royaume-Uni, Oxford University Press.
- Dilhac, M.-A., C. Abrassart, et N. Voarino (éd.), 2018. *La déclaration de Montréal pour un développement responsable de l'intelligence artificielle 2018*. Montréal, QC, Université de Montréal.
- Dodds, L. et P. Wells, 2019. « Data Infrastructure », dans Davies, T., S. B. Walker, M. Rubinstein et F. Perini (éd.), *The State of Open Data*, Cape Town, Afrique du Sud, African Minds et le Centre de recherches pour le développement international (CRDI).

- Données Québec, 2021. Organisations. Adresse : <https://www.donneesquebec.ca/recherche/organization> (consulté en juillet 2021).
- DORA – San Francisco Declaration on Research Assessment, s.d. San Francisco Declaration on Research Assessment. Adresse : <https://sfdora.org/read/> (consulté en septembre 2021).
- Douglas, H., 2015. « Values in Science », dans Humphreys, P. (réd.), *The Oxford Handbook of Philosophy of Science*, Londres, Royaume-Uni, Oxford University Press.
- Dressel, J. et H. Farid, 2018. « The accuracy, fairness, and limits of predicting recidivism », *Science Advances*, vol. 4.
- Egbuonu, K., 2021. The Latest News on the DABUS Patent Case. Adresse : <https://www.ipstars.com/NewsAndAnalysis/the-latest-news-on-the-dabus-patent-case/Index/7366> (consulté en octobre 2021).
- Element AI, 2019. *Rapport mondial 2019 sur les talents en IA*, Montréal QC, Element AI.
- Element AI, 2020. *Rapport mondial sur les talents en IA 2020*, Montréal, QC, Element AI.
- Engler, A., 2021. *How Open-Source Software Shapes AI Policy*, Washington, DC, Brookings Institution.
- EOSC – European Open Science Cloud, 2021. *Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)*, Bruxelles, Belgique, EOSC.
- Eren Vural, I., M. Herder, et J. E. Graham, 2021. « From sandbox to pandemic: Agile reform of Canadian drug regulation », *Health Policy*, vol. 125, n°9, p. 1115-1120.
- Extance, A., 2021. « Machine Learning Delivers ‘Human Genome’ Moment for Proteins ». *Chemistry World* (3 août).
- Ezer, D. et K. Whitaker, 2019. « Data science for the scientific life cycle », *Elife*, vol. 8, p. e43979.
- Fehr, C., 2011. « What Is in It for Me? The Benefits of Diversity in Scientific Communities », dans *Feminist Epistemology and Philosophy of Science*.
- Fiddler, F. et J. Wilcox, 2018. « Reproducibility of Scientific Results », dans Zalta, E. N. (réd.), *Stanford Encyclopedia of Philosophy*, Summer 2021 éd.
- Fife, R. et S. Chase, 2021. « Ottawa Imposes National Security Risk Assessments for University Researchers Seeking Federal Funds ». *The Globe and Mail* (12 juillet).
- Fjeld, J., N. Achten, H. Hilligoss, A. Nagy, et M. Srikumar, 2020. *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, Cambridge (MA), Berkman Klein Center for Internet & Society.
- Fleming, N., 2018. « Computer-calculated compounds », *Nature*, vol. 557, p. S55-S57.
- FLI – Future of Life Institute, 2017a. Asilomar AI Principles. Adresse : <https://futureoflife.org/ai-principles/> (consulté en septembre 2020).

- FLI – Future of Life Institute, 2017b. A Principled AI Discussion in Asilomar.
Adresse : <https://futureoflife.org/2017/01/17/principled-ai-discussion-asilomar/>
(consulté en septembre 2021).
- Floridi, L., 2021. « The European Legislation on AI: A Brief Analysis of its Philosophical Approach », *Philosophy & Technology*.
- Flynn, S., C. Geiger, J. Quintais, T. Margoni, M. Sag, L. Guibault, et M. W. Carroll, 2020. « Implementing User Rights for Research in the Field of Artificial Intelligence », *European Intellectual Property Review*, vol., n°7.
- Fortin, J. M. et D. J. Currie, 2013. « Big science vs. little science: How scientific impact scales with funding », *PLoS One*, vol. 8, n°6, p. e65263.
- Friesen, J., 2021. « CSIS Warns Canadian Universities to be on Alert for International Espionage ». *The Globe and Mail* (26 juillet).
- Fruchterman, J. et J. Mellea, 2018. *Expanding Employment Success for People with Disabilities*, Benetech.
- Gallini, N. et A. Hollis, 2019. *To Sell or Scale Up: Canada's Patent Strategy in a Knowledge Economy*, Montréal, QC, Institute for Research on Public Policy.
- Gaudry, K. et L. Vandsburger, 2020. Determining the Likelihood that an AI patent application will be allowed at the USPTO. Adresse : <https://www.ipwatchdog.com/2020/10/25/determining-likelihood-ai-patent-application-will-allowed-uspto/id=126687/>
(consulté en juillet 2021).
- GC – Gouvernement du Canada, 1985a. *Loi sur les brevets*. R.S.C., 1985, c. P-4, Ottawa, ON, GC.
- GC – Gouvernement du Canada, 2020b. Infrastructure canadienne de données géospatiales. Adresse : <https://www.rncan.gc.ca/science-et-donnees/science-et-recherche/sciences-terre/geomatique/infrastructure-canadienne-donnees-geospatiales/18771>
(consulté en août 2021).
- GC – Gouvernement du Canada, 1985b. *Loi sur l'accès à l'information*. R.S., 1985 c. A-1, s. 20, Ottawa, ON, GC.
- GC – Gouvernement du Canada, 2019. Stratégie en matière de propriété intellectuelle. Adresse : <https://www.ic.gc.ca/eic/site/108.nsf/fra/accueil> (consulté en avril 2021).
- GC – Gouvernement du Canada, 2020a. Open Data 101. Adresse : <https://open.canada.ca/en/open-data-principles> (consulté en décembre 2021).
- GC – Gouvernement du Canada, 2020c. *Rapport annuel de la conseillère scientifique en chef 2019-20*, Ottawa, ON, Bureau de la conseillère scientifique en chef.

- GC – Gouvernement du Canada, 2021a. Le gouvernement du Canada lance une consultation sur un cadre moderne du droit d'auteur pour l'intelligence artificielle et l'Internet des objets. Adresse : <https://www.canada.ca/fr/innovation-sciences-developpement-economique/nouvelles/2021/07/le-gouvernement-du-canada-lance-une-consultation-sur-un-cadre-moderne-du-droit-dauteur-pour-lintelligence-artificielle-et-linternet-des-objets.html> (consulté en juillet 2021).
- GC – Gouvernement du Canada, 2021b. Données ouvertes. Adresse : <https://ouvert.canada.ca/fr/donnees-ouvertes> (consulté en juillet 2021).
- GC – Gouvernement du Canada, 2021c. *Budget 2021 — Une relance axée sur les emplois, la croissance et la résilience*, Ottawa, ON, Ministère des Finances Canada.
- GC – Gouvernement du Canada, 2021d. Politique des trois organismes sur la gestion des données de recherche. Adresse : https://science.gc.ca/eic/site/063.nsf/fra/h_97610.html (consulté en juillet 2021).
- GC – Gouvernement du Canada, 2021e. Modèle de politique sur l'intégrité scientifique. Adresse : https://www.ic.gc.ca/eic/site/063.nsf/fra/h_97643.html (consulté en septembre 2021).
- GC – Gouvernement du Canada, 2021f. Directive sur la prise de décisions automatisée. Adresse : <https://www.tbs-sct.gc.ca/pol/doc-fra.aspx?id=32592§ion=html> (consulté en mai 2021).
- GC – Gouvernement du Canada, 2021g. Un cadre de mise en œuvre de l'ouverture par défaut de la recherche scientifique fédérale. Adresse : https://science.gc.ca/eic/site/063.nsf/fra/h_98201.html (consulté en juillet 2021).
- GDPR – General Data Protection Regulation, 2021. General Data Protection Regulation (GDPR). Adresse : <https://gdpr.eu/article-89-processing-for-archiving-purposes-scientific-or-historical-research-purposes-or-statistical-purposes/> (consulté en août 2021).
- Gebru, T., J. Morgenstern, B. Vecchione, J. Wortman Vaughan, H. Wallach, H. Daumé III, et K. Crawford, 2021. « Datasheets for datasets », *arXiv*, n°1803.09010v8.
- GEHN IA – Groupe d'experts de haut niveau sur l'intelligence artificielle, 2019. *Lignes directrices en matière d'éthique pour une IA digne de confiance*, Bruxelles, Belgique, Commission européenne.
- Geist, M., 2021. « AI and International Regulation », dans Martin-Bariteau, F. et T. Scassa (réd.), *Artificial Intelligence and the Law in Canada*, Toronto, ON, LexisNexis.
- Gherhes, C., T. Vorley, P. Vallance, et C. Brooks, 2021. « The role of system-building agency in regional path creation: Insights from the emergence of artificial intelligence in Montreal », *Regional Studies*, vol., p. 1-16.
- Gibert, M., 2020. *Faire la morale aux robots*. Montreal, QC, Atelier 10.
- Gibert, M., C. Mondin, et G. Chicoisne, 2018. « Partie 2 : Portrait 2018 des recommandations internationales en éthique de l'IA », dans Dilhac, M.-A., C. Abrassart et N. Voarino (réd.), *La Déclaration de Montréal IA responsable*, Montréal, QC, Université de Montréal.

- GIDA – Global Indigenous Data Alliance, s.d.-a. Purpose. Adresse : <https://www.gida-global.org/purpose> (consulté en juin 2021).
- GIDA – Global Indigenous Data Alliance, s.d.-b. Care Principles for Indigenous Data Governance. Adresse : <https://www.gida-global.org/care> (consulté en juin 2021).
- Gil, Y., M. Greaves, J. Hendler, et H. Hirsh, 2014. « Amplify scientific discovery with artificial intelligence », *Science*, vol. 346, n°6206, p. 171-172.
- Gil, Y., R. D. King, et H. Kitano, 2020. *AI Scientist Grand Challenge — Summary of Discussion During Workshop Held in February 2020*, Londres, Royaume-Uni, Alan Turing Institute.
- Giuffrida, I., 2019. « Liability for AI decision-making: Some legal and ethical considerations », *Fordham Law Review*, vol. 88, n°2, p. 439-456.
- Gouv. de l'Ont. – Gouvernement de l'Ontario, 2017. L'Ontario nomme sa toute première directrice du numérique. Adresse : <https://news.ontario.ca/fr/release/4,4109/ontario-nomme-sa-toute-premiere-directrice-du-numerique> (consulté en septembre 2021).
- Gouv. de l'Ont. – Gouvernement de l'Ontario, 2021a. *Modernisation de la protection de la vie privée en Ontario*. Toronto, ON, Gouv. de l'ON.
- Gouv. de l'Ont. – Gouvernement de l'Ontario, 2021c. Conseils sur l'Intelligence Artificielle (IA). Adresse : <https://www.ontario.ca/fr/page/conseils-sur-lintelligence-artificielle-ia> (consulté en mai 2021).
- Gouv. de l'Ont. – Gouvernement de l'Ontario, 2021b. Consultation : Office ontarien des données. Adresse : <https://www.ontario.ca/fr/page/consultation-office-ontarien-des-donnees> (consulté en septembre 2021).
- Gouv. du Royaume-Uni – Gouvernement du Royaume-Uni, 2020. National Data Strategy. Adresse : <https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy> (consulté en juillet 2021).
- Gouvernement de l'Australie, 2021. National Data Commissioner. Adresse : <https://pmc.gov.au/public-data/national-data-commissioner> (consulté en août 2021).
- Gouvernement de la Nouvelle-Zélande, 2020. Government Chief Data Steward. Adresse : <https://www.digital.govt.nz/digital-government/leadership/government-functional-leads/government-chief-data-steward-gclds/> (consulté en août 2021).
- Grant, J., 2017. « The Allocation of Scientific Grants Should be a Science ». *Times Higher Education* (15 juin).
- GRC – Global Research Council, 2021. *Responsible Research Assessment — Global Research Council (GRC) Conference Report 2021*, GRC.
- Gundersen, O. E. et S. Kjensmo, 2018. *State of the Art: Reproducibility in Artificial Intelligence*, communication présentée dans le cadre du The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18).

- Gundersen, O. E., Y. Gil, et D. W. Aha, 2018. « On reproducible AI: towards reproducible research, open science, and digital scholarship in AI publications », *AI Magazine*, vol. 39, n°3, p. 56-68.
- Gunning, D. et D. W. Aha, 2017. « DARPA's explainable artificial intelligence (XAI) program », *AI Magazine*, vol. 40, n°2, p. 44-58.
- Guo, J., V. Farhang-Razi, et P. R. Algra, 2019. « AI: A Glossary of Terms », dans Erik R. Ranschaert, E. R., S. Sergey Morozov et P. R. Algra (réd.), *Artificial Intelligence in Medical Imaging*, 1 éd., Cham, Suisse, Springer.
- Gupta, A., C. Lanteigne, et S. Kingsley, 2020. *A Social and Environmental Certificate for AI Systems*, Montréal, QC, Montreal AI Ethics Institute.
- Habibollahi Najaf Abadi, H. et M. Pecht, 2020. « Artificial Intelligence Trends Based on the Patents Granted by the United States Patent and Trademark Office », *IEEE Access*, vol. 8, p. 81633-81643.
- Haenlein, M. et A. Kaplan, 2019. « A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence », *California Management Review*, vol. 61, n°4, p. 5-14.
- Hagen, G. D., 2021. « AI and Patents and Trade Secrets », dans Martin-Bariteau, F. et T. Scassa (réd.), *Artificial Intelligence and the Law in Canada*, Toronto, ON, LexisNexis.
- HAI – Stanford University Human-Centered Artificial Intelligence, 2021. *Artificial Intelligence Index Report 2021*, Stanford, CA, Stanford University Human-Centered Artificial Intelligence.
- Haibe-Kains, B., G. A. Adam, A. Hosny, F. Khodakarami, Massive Analysis Quality Control (MAQC) Society Board of Directors, L. Waldron, ... H. J. W. L. Aerts, 2020. « Transparency and reproducibility in artificial intelligence », *Nature*, vol. 586, p. E14-E16.
- Hajian, S., F. Bonchi, et C. Castillo, 2016. *Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining*, communication présentée dans le cadre du 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA.
- Hatfield, P. W., J. A. Gaffney, G. J. Anderson, S. Ali, L. Antonelli, S. Basegmez du Pree, ... B. Williams, 2021. « The data-driven future of high-energy-density physics », *Nature*, vol. 593, n°7859, p. 351-361.
- Heaven, W. D., 2018. « AI peer reviewers unleashed to ease publishing grind ». *Nature News* (22 novembre).
- Heaven, W. D., 2020. « AI is Wrestling with a Replication Crisis ». *MIT Technology Review* (12 novembre).
- Hecht, B., L. Wilcox, J. P. Bigham, J. Schöning, E. Hoque, J. Ernst, ... C. Wu, 2018. It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process. Adresse : <https://acm-fca.org/2018/03/29/negativeimpacts/> (consulté en juin 2021).

- Hellström, T., V. Dignum, et S. Bensch, 2020. « Bias in machine learning — what is it good for? », *arXiv*, n°2004.00686v2.
- Helmers, L., F. Horn, F. Biegler, T. Oppermann, et K. R. Muller, 2019. « Automating the search for a patent's prior art with a full text similarity search », *PLoS One*, vol. 14, n°3, p. e0212103.
- Henderson, P., J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, et J. Pineau, 2020. « Towards the systematic reporting of the energy and carbon footprints of machine learning », *arXiv*.
- Henderson, P., R. Islam, P. Bachman, J. Pineau, D. Precup, et D. Meger, 2018. *Deep Reinforcement Learning that Matters*, communication présentée dans le cadre du The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA.
- Hervey, M., 2020. « Harnessing AI in drug discovery without losing patent protection », *Drug Discovery Today*, vol. 25, n°6, p. 949-950.
- Hinton, J. W. et P. Cowan, 2018. « Canada's IP Strategy is Not in Step with Our Innovation and Commercialization Goals ». *The Globe and Mail* (25 novembre).
- Hinton, J. W., 2020. « Canada Needs to Own Critical IP and Data Assets to Inspire Generational Economic Prosperity ». *The Hill Times* (25 mars).
- Hocky, G. M. et A. D. White, 2022. « Natural language processing models that automate programming will transform chemistry research and teaching », *arXiv*, n°2108.13360v2.
- Hofstra, B., V. V. Kulkarni, S. Munoz-Najar Galvez, B. He, D. Jurafsky, et D. A. McFarland, 2020. « The Diversity-innovation paradox in science », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, n°17, p. 9284-9291.
- Hou, M., G. Ho, et D. Dunwoody, 2021. « IMPACTS: A trust model for human-autonomy teaming », *Human-Intelligent Systems Integration*, vol. 3, n°2, p. 79-97.
- Humphreys, P., 2009. « The philosophical novelty of computer simulation methods », *Synthese*, vol. 169, p. 615-626.
- Hutchinson, B., V. Prabhakaran, E. Denton, K. Webster, Y. Zhong, et S. Denuyl, 2020. « Unintended Machine Learning Biases as Social Barriers for Persons with Disabilities », *ACM SIGACCESS Accessibility and Computing*.
- Hutson, M., 2018. « Missing Data Hinder Replication of Artificial Intelligence Studies ». *Science Magazine* (15 février).
- Hutson, M., 2021. « AI conferences use AI to assign papers to reviewers », *Science Insider* (1 avril)
- INDU – Comité permanent de l'industrie et de la technologie, 2019. *Examen prévu par la loi de la loi sur le droit d'auteur*, Ottawa, ON, Chambre des Communes.
- Institut Vecteur, s.d. *Intellectual Property Policy*, Toronto, ON, Institut Vecteur.
- IPWatchdog, 2021. DABUS Gets its first Patent in South Africa under Formalities Examination. Adresse : <https://www.ipwatchdog.com/2021/07/29/dabus-gets-first-patent-south-africa-formalities-examination/id=136116/> (consulté en juillet 2021).

- IRSC – Instituts de recherche en santé du Canada, 2019. Politique des trois organismes sur le libre accès aux publications. Adresse : <https://cihr-irsc.gc.ca/f/32005.html> (consulté en juillet 2021).
- IRSC – Instituts de recherche en santé du Canada, 2021. Aperçu du financement. Adresse : <https://cihr-irsc.gc.ca/f/37788.html> (consulté en février 2022).
- IRSC, CRSNG, et CRSH – Instituts de recherche en santé du Canada, Conseil de recherches en sciences naturelles et en génie du Canada, Conseil de recherches en sciences humaines du Canada, 2016. *Cadre de référence des trois organismes sur la Conduite responsable de la recherche*, Ottawa, ON, IRSC, CRSNG, et CRSH.
- IRSC, CRSNG, et CRSH – Instituts de recherche en santé du Canada, Conseil de recherches en sciences naturelles et en génie du Canada, Conseil de recherches en sciences humaines du Canada, 2018a. *Plan d'action des trois organismes pour l'équité, la diversité et l'inclusion (2018-2025)*, Ottawa, ON, IRSC, CRSNG, et CRSH.
- IRSC, CRSNG, et CRSH – Instituts de recherche en santé du Canada, Conseil de recherches en sciences naturelles et en génie du Canada, Conseil de recherches en sciences humaines du Canada, 2018b. *Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains*, Ottawa, ON, IRSC, CRSNG, et CRSH.
- Isdahl, R. et O. E. Gundersen, 2019. *Out-of-the-box Reproducibility: A Survey of Machine Learning Platforms*, communication présentée dans le cadre du 2019 IEEE 15th International Conference on e-Science and Grid Computing, San Diego, CA.
- ISDE – Innovation, Sciences et Développement économique Canada, 2021a. *Consultation sur un cadre moderne du droit d'auteur pour l'intelligence artificielle et l'Internet des objets*, Ottawa, ON, ISDE.
- ISDE – Innovation, Sciences et Développement économique Canada, 2021c. Lignes directrices sur l'examen relatif à la sécurité nationale des investissements. Adresse : <https://www.ic.gc.ca/eic/site/ica-lic.nsf/fra/lk81190.html> (consulté en avril 2021).
- ISDE – Innovation, Sciences et Développement économique Canada, 2019a. *Traitement de l'intelligence artificielle : Aperçu du paysage canadien des brevets*, Ottawa, ON, Government of Canada.
- ISDE – Innovation, Sciences et Développement économique Canada, 2019b. *La Charte numérique du Canada en action : un plan par des Canadiens, pour les Canadiens*, Ottawa, ON, Gouvernement du Canada.
- ISDE – Innovation, Sciences et Développement économique Canada, 2021b. Le gouvernement du Canada agit pour protéger la recherche et la propriété intellectuelle au Canada. Adresse : <https://www.canada.ca/fr/innovation-sciences-developpement-economique/nouvelles/2021/07/le-gouvernement-du-canada-agit-pour-protger-la-recherche-et-la-proprie-intellectuelle-au-canada.html> (consulté en août 2021).

- ITK – Inuit Tapiriit Kanatami, s.d. National Inuit Strategy on Research. Adresse : <https://www.itk.ca/national-strategy-on-research-launched/> (consulté en août 2021).
- Jia, X., A. Lynch, Y. Huang, M. Danielson, I. Lang'at, A. Milder, ... J. Schrier, 2019. « Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis », *Nature*, vol. 573, n°7773, p. 251-255.
- Jobin, A., M. Ienca, et E. Vayena, 2019. « The Global Landscape of AI Ethics Guidelines », *Nature Machine Intelligence*, vol. 1, n°389, p. 389-399.
- Johnson, L., 2021. Canada's AI Mafia. Adresse : <https://xtract.ai/the-history-of-ai/> (consulté en mai 2021).
- Johnson, P. A., R. Sieber, T. Scassa, M. Stephens, et P. Robinson, 2017. « The cost(s) of geospatial open data », *Transactions in GIS*, vol. 21, n°3, p. 434-445.
- JOURNE, 2021. JOURNE — Journal of Opportunities, Unexpected limitations, Retrospectives, Negative results, and Experiences. Adresse : <https://sites.google.com/view/workshop-journe/home> (consulté en août 2021).
- Kalvapalle, R., 2021. « U of T Acceleration Consortium to use AI to develop advanced materials ». *U of T News* (20 avril).
- Kelly, J., 2016. The text and data mining copyright exception - benefits and implications for UK higher education. Adresse : <https://www.jisc.ac.uk/guides/text-and-data-mining-copyright-exception> (consulté en août 2021).
- King, R. D. et S. Roberts, 2018. « Chapter 5. Artificial Intelligence and Machine Learning in Science », dans OCDE (réd.), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, Paris, France, les Editions de l'OCDE.
- King, R. D., J. Rowland, S. G. Oliver, M. Young, W. Aubrey, E. Byrne, ... A. Clare, 2009. « The automation of science », *Science*, vol. 324, n°5923, p. 85-89.
- King, R. D., V. Schüler-Costa, C. Mellingwood, et L. Soldatova, 2018. « Automating sciences: Philosophical and social dimensions », *IEEE Technology and Society Magazine*, vol. 37, n°1, p. 40-46.
- Kitano, H., 2021. « Nobel Turing Challenge: Creating the engine for scientific discovery », *NPJ Systems Biology and Applications*, vol. 7, n°1, p. 29.
- Knight, W., 2017. « The Dark Secret at the Heart of AI ». *MIT Technology Review* (11 avril).
- Knubley, J., 2021. *Building Superclusters for Canada*, Toronto, ON, Brookfield.
- Kramer, A. D. I., J. E. Guillory, et J. T. Hancock, 2014. « Experimental evidence of massive-scale emotional contagion through social networks », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, n°24, p. 8788-8790.
- Kriegman, S., D. Blackiston, M. Levin, et J. Bongard, 2020. « A scalable pipeline for designing reconfigurable organisms », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, n°4, p. 1853-1859.

- Kung, J., G. Boskovic, et C. Stix, 2020. *Building an AI World: Report on National and Regional AI Strategies*, Toronto, ON, Canadian Institute for Advanced Research.
- Kusters, R., D. Misevic, H. Berry, A. Cully, Y. Le Cunff, L. Dandoy, ... F. Wehbi, 2020. « Interdisciplinary research in artificial Intelligence: Challenges and opportunities », *Frontiers in Big Data*, vol. 3, p. 577974.
- L'Alliance – Alliance de recherche numérique du Canada, 2021. Dévoilement de l'Alliance de recherche numérique du Canada. Adresse : <https://alliancecan.ca/fr/nouveautes/nouvelles/devoilement-de-lalliance-de-recherche-numerique-du-canada> (consulté en octobre 2021).
- Lacoste, A., A. Luccioni, V. Schmidt, et T. Dandres, 2019. « Quantifying the carbon emissions of machine learning [pre-print] », *arXiv*, n°1910.09700v2.
- Lamb, C. et D. Munro, 2020. *The Intangible Shift: Changing Gears to Compete in the New Economy*, Toronto, ON, The Brookfield Institute for Innovation + Entrepreneurship.
- Lane, M. et A. Saint-Martin, 2021. *The Impact of Artificial Intelligence on the Labour Market: What do We Know so Far?*, Paris, France, OCDE.
- Lapuschkin, S., S. Wäldchen, A. Binder, G. Montavon, W. Samek, et K.-R. Müller, 2019. « Unmasking Clever Hans predictors and assessing what machines really learn », *Nature Communications*, vol. 10, p. 1096.
- Larivière, V. et C. Sugimoto, 2018. « Do authors comply with mandates for open access? », *Nature*, vol. 562, p. 483-486.
- Laudel, G., 2006. « The 'Quality Myth': Promoting and hindering conditions for acquiring research funds », *Higher Education*, vol. 52, p. 375-403.
- Leblond, P., 2019. *Digital Trade at the WTO — The CPTPP and CUSMA Pose Challenges to Canadian Data Regulation*, Waterloo, ON, Centre for International Governance Innovation.
- Lehr, D. et P. Ohm, 2017. « Playing with the data: What legal scholars should learn about machine learning », *UC Davis Law Review*, vol. 51, p. 653-717.
- Leonelli, S., 2018. « Rethinking Reproducibility as a Criterion for Research Quality », dans Fiorito, S. L., S. Scheall et C. Eduardo (réd.), *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise*, vol. 36B, Bingley, Royaume-Uni, Emerald Publishing Limited.
- Leonelli, S., 2020. « Scientific Research and Big Data », dans Zalta, E. N. (réd.), *The Stanford Encyclopedia of Philosophy*, Summer 2020 éd.
- Leonelli, S., R. Lovell, B. W. Wheeler, L. Fleming, et H. Williams, 2021. « From FAIR data to fair data use: Methodological data fairness in health-related social media research », *Big Data & Society*, vol. 8, n°1.
- Lewis, J. E., 2020. *Indigenous Protocol and Artificial Intelligence Position Paper*, Honolulu, HI, The Initiative for Indigenous Futures et l'Institut canadien de recherches avancées (CIFAR).
- Lin, D., J. Crabtree, I. Dillo, R. R. Downs, R. Edmunds, D. Giarretta, ... J. Westbrook, 2020. « The TRUST Principles for digital repositories », *Scientific Data*, vol. 7, n°1, p. 144.

- Lipton, Z. C., 2018. « The mythos of model interpretability », *Queue*, vol. 16, n°3, p. 31-57.
- Lo, B. et M. J. Field (réd.), 2009. *Conflict of Interest in Medical Research, Education, and Practice*. Washington, DC, The National Academies Press.
- Local Contexts, s.d. About Us. Adresse : <https://localcontexts.org/about/about-local-contexts/> (consulté en juin 2021).
- Logg, J. M., J. A. Minson, et D. A. Moore, 2018. « Algorithm appreciation: People prefer algorithmic to human judgment », *Organizational Behavior and Human Decision Processes*, vol. 151, p. 90-103.
- Lopez, C. E., S. R. Miller, et C. S. Tucker, 2019. « Exploring biases between human and machine generated designs », *Journal of Mechanical Design*, vol. 141, n°2.
- Lopez-Gonzalez, J., F. Casalini, et T. Nemoto, 2021. « Mapping Approaches to Cross-Border Data Flows », dans Borchert, I. et L. A. Winters (réd.), *Addressing Impediments to Digital Trade*, CEPR Press.
- Lorsch, J. R., 2015. « Maximizing the return on taxpayers' investments in fundamental biomedical research », *Molecular Biology of the Cell*, vol. 26, n°9, p. 1578-1582.
- Lowrie, I., 2017. « Algorithmic rationality: Epistemology and efficiency in the data sciences », *Big Data & Society*, vol. 4, n°1.
- MacGregor, I., 2018. « Big Data - The Canadian Opportunity ». *CIGIOnline* (18 mars).
- Malone, M., 2020. « Trade secrets, big data, and the future of public interest litigation over AI in Canada », *Canadian Intellectual Property Review*, vol. 35, p. 6-9.
- Marcia, V. et K. C. DeSouza, 2021. « The EU Path Towards Regulation on AI ». *Brookings TechTank* (26 avril).
- McKelvey, F. et J. Roberge, 2021. « Canada is gambling with its leadership on artificial intelligence ». *The Globe and Mail* (avril 25).
- Medeiros, M., W. Chalmers, et J. Sanft, 2021. « Intellectual Property Strategy for AI Technology in Canada », dans D'Agostino, G., A. Gaon et C. Piovesan (réd.), *Leading Legal Disruption: Artificial Intelligence and a Tool Kit for Lawyers and the Law*, Toronto, ON, Thomson Reuters Canada.
- Mercurio, L. et H. Y. Yu, 2021. « An AI Policy for the (Near) Future », dans Borchert, I. et L. A. Winters (réd.), *Addressing Impediments to Digital Trade*, CEPR Press.
- Merton, R., 1968. « The Matthew Effect in Science », *Science*, vol. 159, n°3810, p. 56-63.
- Metcalf, J. et K. Crawford, 2016. « Where are human subjects in big data research? The emerging ethics divide », *Big Data & Society*, vol. 3, n°1.
- Metcalf, J., E. F. Keller, et d. boyd, 2021. *Perspectives on Big Data, Ethics, and Society*, The Council for Big Data, Ethics, and Society.
- Meyer, M. N., 2014. « Misjudgements will drive social trials underground », *Nature*, vol. 511, p. 265.

- MGCS – Ministère des Services gouvernementaux et des Services aux consommateurs, 2021. *Manuel sur l'accès à l'information et la protection de la vie privée*, Ottawa, ON, Gouvernement de l'Ontario.
- Mila – Mila-Quebec Artificial Intelligence Institute, 2021. Politique de partenariat. Adresse : <https://mila.quebec/mila/partenaires-industriels/police-partenariat/> (consulté en juillet 2021).
- Mongeon, P., C. Brodeur, C. Beaudry, et V. Larivière, 2016. « Concentration of research funding leads to decreasing marginal returns », *Research Evaluation*, vol. 25, n°4, p. 396-404.
- Mons, B., C. Neylon, J. Velterop, M. Dumontier, L. O. B. da Silva Santos, et M. D. Wilkinson, 2017. « Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud », *Information Services & Use*, vol. 37, n°1, p. 49-56.
- Morley, J., L. Floridi, L. Kinsey, et A. Elhalal, 2021. « From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices », dans L., F. (réd.), *Ethics, Governance, and Policies in Artificial Intelligence. Philosophical Studies Series*, vol. 144, Cham, Suisse, Springer.
- Morrison, O., 2019. Invention In the Age of Artificial Intelligence. What About Us? Adresse : <https://medium.com/swlh/invention-in-the-age-of-artificial-intelligence-what-about-us-6fa4451c39a2> (consulté en septembre 2021).
- Mourby, M., H. Gowans, S. Aidinlis, H. Smith, et J. Kaye, 2019. « Governance of academic research data under the GDPR — Lessons from the UK », *International Data Privacy Law*, vol. 9, n°3, p. 192-206.
- Mugleton, S. H., U. Schmid, C. Zeller, A. Tamaddoni-Nezhad, et T. Besold, 2018. « Ultra-Strong Machine Learning: Comprehensibility of programs learned with ILP », *Machine Learning*, vol. 107, n°7, p. 1119-1140.
- Muller, V. C., 2020. « Ethics of Artificial Intelligence and Robotics », dans Zalta, E. N. (réd.), *The Stanford Encyclopedia of Philosophy*, Winter 2020 éd.,
- Mulligan, C. et S. Elaluf-Calderwood, 2021. « AI ethics: A framework for measuring embodied carbon in AI systems », *AI and Ethics*.
- Mullin, R., 2021. « The Lab of the Future is Now ». *Chemical & Engineering News* (28 mars).
- Nagarajan, P., G. Warnell, et P. Stone, 2019. « Deterministic implementations for reproducibility in deep reinforcement learning », *arXiv*, n°1809.05676v5.
- NAIO – National Artificial Intelligence Initiative Office, 2021. Legislation and Executive Orders. Adresse : <https://www.ai.gov/legislation-and-executive-orders/> (consulté en septembre 2021).
- NASEM – National Academies of Sciences, Engineering, and Medicine, 2018. *Data Science: Opportunities to Transform Chemical Sciences and Engineering: Processing of a Workshop in Brief*, Washington, DC, NASEM.

- NASEM – National Academies of Sciences, Engineering, and Medicine, 2019. *Reproducibility and Replicability in Science*. Washington, DC, The National Academies Press.
- NASEM – National Academies of Sciences, Engineering, and Medicine, 2020. *Information Technology Innovation: Resurgence, Confluence, and Continuing Impact*. Washington, DC, The National Academies Press.
- Naylor, C. D., R. J. Birgeneau, M. Crago, M. Lazaridis, C. Malacrida, A. B. McDonald, ... A. Wilson, 2017. *Investir dans l'avenir du Canada : Consolider les bases de la recherche au pays*, Ottawa, ON, L'examen du soutien fédéral aux sciences.
- NDRIO – New Digital Research Infrastructure Organization, 2021a. *Working Together to Create a Truly Integrated DRI Ecosystem in Canada*, communication présentée dans le cadre du 4^e National Data Services Framework Summit, Virtual Summit.
- NDRIO – New Digital Research Infrastructure Organization, 2021b. *Canadian DRI Needs Assessment - Preliminary Findings*, Toronto, ON, NDRIO.
- NeurIPS – Conference on Neural Information Processing Systems, 2020. NeurIPS 2020 FAQ for Authors. Adresse : <https://nips.cc/Conferences/2020/PaperInformation/NeurIPS-FAQ> (consulté en juin 2021).
- NeurIPS – Conference on Neural Information Processing Systems, 2021. NeurIPS 2021 Paper Checklist Guidelines. Adresse : <https://neurips.cc/Conferences/2021/PaperInformation/PaperChecklist> (consulté en juin 2021).
- Nickles, T., 2018. « Alien reasoning: Is a major change in scientific research underway? », *Topoi*, vol. 39, n°4, p. 901-914.
- Niczepyr, D., 2021. « National AI Initiative Office Launched by White House ». *Fedscoop* (12 janvier).
- Nielsen, M. W., S. Alegria, L. Börjeson, H. Etzkowitz, H. J. Falk-Krzesinski, A. Joshi, ... L. Schiebinger, 2017. « Opinion: Gender diversity leads to better science », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, n°13, p. E2796.
- NIST – National Institute of Standards and Technology, 2020. High-Throughput Experimental Materials Collaboratory. Adresse : <https://www.nist.gov/programs-projects/high-throughput-experimental-materials-collaboratory> (consulté en mai 2021).
- Obermeyer, Z., B. Powers, C. Vogeli, et S. Mullainathan, 2019. « Dissecting racial bias in an algorithm used to manage the health of populations », *Science*, vol. 366, p. 477-453.
- OCDE – OECD.AI Policy Observatory, 2021e. *OECD Metrics & Methods*, Paris, France, les Editions de l'OCDE.
- OCDE – Organisation de Coopération et de Développement Économiques, 2018a. *Intelligence Machines, Smart Policies*, communication présentée dans le cadre du AI: Intelligent Machines, Smart Policies, Paris, France.

- OCDE – Organisation de Coopération et de Développement Économiques, 2018b. *Science, technologie et innovation : Perspectives de l'OCDE 2021 : Affronter la crise et saisir les opportunités*, Paris, France, les Editions de l'OCDE.
- OCDE – Organisation de Coopération et de Développement Économiques, 2019a. Recommandation du Conseil sur l'intelligence artificielle. Adresse : <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449> (consulté en septembre 2021).
- OCDE – Organisation de Coopération et de Développement Économiques, 2019b. *Scoping the OECD AI Principles — Deliberations of the Expert Group on Artificial Intelligence at the OECD*, Paris, France, les Editions de l'OCDE.
- OCDE – Organisation de Coopération et de Développement Économiques, 2021a. *Tools for Trustworthy AI*, Paris, France, les Editions de l'OCDE.
- OCDE – Organisation de Coopération et de Développement Économiques, 2021b. *State of Implementation of the OECD AI Principles — Insights from National AI Policies*, Paris, France, les Editions de l'OCDE.
- OCDE – Organisation de Coopération et de Développement Économiques, 2021c. *Recommandation du Conseil sur l'intelligence artificielle*. Paris, France, les Editions de l'OCDE.
- OCDE – Organisation de Coopération et de Développement Économiques, 2021d. Shaping an enabling policy environment for AI (Principe 2.3). Adresse : <https://www.oecd.ai/dashboards/ai-principles/P12> (consulté en août 2021).
- O'Connor, A. M., G. Tsafnat, J. Thomas, P. Glasziou, S. B. Gilbert, et B. Hutton, 2019. « A question of trust: Can we build an evidence base to gain trust in systematic review automation technologies? », *Systematic Reviews*, vol. 8, n°1, p. 143.
- OMPI – Organisation Mondiale de la Propriété Intellectuelle, 2020. *Dialogue de l'OMPI sur la propriété intellectuelle Et l'intelligence artificielle*, Genève, Suisse, OMPI.
- OPIC – Office de la propriété intellectuelle du Canada, 2021. Qu'est-ce qu'un brevet? Adresse : <https://www.ic.gc.ca/eic/site/cipointernet-internetopic.nsf/fra/wr03716.html> (consulté en juin 2021).
- OPIC – Office de la propriété intellectuelle du Canada,, 2019. *Rapport sur la PI au Canada*, Ottawa, ON, Innovation, Sciences et Développement économique Canada.
- Oransky, I. et A. Marcus, 2017. « Peer Review Is A Black Box. Let's Open It Up ». *Stat* (20 juillet).
- Parlement du Canada, 2020. *Loi portant mise en œuvre de l'Accord entre le Canada, les États-Unis d'Amérique et les États-Unis mexicains*, Ottawa, ON, Parlement du Canada.
- Parlement européen, 2019a. *Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE*, Bruxelles, Belgique, Parlement européen.

- Parlement européen, 2019b. *Directive (EU) 2019/1024 of the European Parliament and the Council*. Bruxelles, Belgique, Parlement européen.
- Parlement européen, 2020. Résolution du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission sur un régime de responsabilité civile pour l'intelligence artificielle (2020/2014 (INL)). Adresse : https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_FR.html (consulté en juillet 2021).
- Partnership on AI, 2021. *Managing the Risks of AI Research – Six Recommendations for Responsible Publication*, San Francisco, CA, Partnership on AI.
- Patterson, D., J. Gonzalez, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, ... J. Dean, 2021. « Carbon Emissions and Large Neural Network Training [pre-print] », arXiv, n°2104.10350v3.
- PODS – La Plateforme ontarienne des données sur la santé, 2020. la PODS. Adresse : <https://ohdp.ca/apercu/?lang=fr> (consulté en août 2021).
- Pritt, M., 2018. Between a Rock and a Hard Place: Protecting Employee Mobility and Negative Trade Secrets. Adresse : <https://www.bsflp.com/news-events/between-a-rock-and-a-hard-place-protecting-employee-mobility-and-negative-trade-secrets.html> (consulté en juillet 2021).
- Rainie, S. C., D. Rodriguez-Lonebear, et A. Martinez, 2017. *Indigenous Data Sovereignty in the United States*, Tuscon, AZ, University of Arizona Natives Nations Institute.
- Reardon, S., 2015. « Canadians balk at reforms to health-research agency », *Nature*, vol. 520, n°7547, p. 272-273.
- Renda, A., 2019. *Artificial Intelligence — Ethics, Governance, and Policy Challenges*, Bruxelles, Belgique, Centre for European Policy Studies (CEPS).
- Reynolds, A. et D. Lewis, 2017. « Teams Solve Problems Faster When They're More Cognitively Diverse ». *Harvard Business Review* (30 mars).
- Reznick, R. K., K. Harris, et T. Horsley, 2020. *Task Force Report on Artificial Intelligence and Emerging Digital Technologies*, Ottawa, ON, Le Collège royal des médecins et chirurgiens du Canada.
- RNCan – Ressources naturelles Canada, 2019. *Canadian Geospatial Data Infrastructure (CGDI) — User Needs Assessment*, Ottawa, ON, RNCan.
- RNCan – Ressources naturelles Canada, 2020. *Cahier d'information sur l'énergie 2020-2021*, Ottawa, ON, RNCan.
- Roberts, H., J. Cowsls, E. Hine, J. Morley, M. Taddeo, V. Wang, et L. Floridi, 2021. « China's artificial intelligence strategy: Lessons from the European Union's 'ethics-first' approach », *Social Science Research Network*.
- Roberts, J. et D. Fisher, 2020. *pReview: The Artificially Intelligent Conference Reviewer*, communication présentée dans le cadre du 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL.

- Rolin, K., 2019. « The Epistemic Significance of Diversity », dans Fricker, M., P. J. Graham, D. Henderson et N. J. L. Pedersen (réd.), *The Routledge Handbook of Social Epistemology*, New York, NY, Routledge.
- RRITools, s.d. *A Practical Guide to Responsible Research and Innovation*, Valencia, Spain, La Caixa Foundation.
- Rudin, C. et J. Radin, 2019. « Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition », *Harvard Data Science Review*, vol., n°1.2.
- Rudin, C., 2019. « Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead », *Nature Machine Intelligence*, vol. 1, n°5, p. 206-215.
- Samek, W. et K.-R. Müller, 2019. « Towards Explainable Artificial Intelligence », *arXiv*, n°1909.12072v1.
- Sanderson, K., 2019. « Automation: Chemistry shoots for the moon », *Nature*, vol. 568, n°577-579.
- Sap, M., D. Card, S. Gabriel, Y. Choi, et N. A. Smith, 2019. *The Risk of Racial Bias in Hate Speech Detection*, communication présentée dans le cadre du 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italie.
- Satariano, A., 2020. « British Grading Debacle Shows Pitfalls of Automating Government ». *The New York Times* (20 août).
- SC – Santé Canada, 2019. *Réglementation agile pour les produits thérapeutiques de pointe et les essais cliniques*, Ottawa, ON, SC.
- SC – Santé Canada, 2021. Innovation réglementaire pour les produits de santé : Favoriser l'utilisation de produits thérapeutiques de pointe. Adresse : <https://www.canada.ca/fr/sante-canada/organisation/a-propos-sante-canada/activites-responsabilites/strategies-initiatives/modernisation-reglementation-aliments-produits-sante/produits-therapeutiques-avances.html> (consulté en octobre 2021).
- Scassa, T., 2019a. « Ownership and control over publicly accessible platform data », *Online Information Review*, vol. 43, n°6, p. 986-1002.
- Scassa, T., 2019b. « Privacy », dans Davies, T., S. B. Walker, M. Rubinstein et F. Perini (réd.), *The State of Open Data*, Cape Town, Afrique du Sud, African Minds et le Centre de recherches pour le développement international (CRDI).
- Scassa, T., 2021a. « Legal Issues Around Data Scraping as a Source of Data for AI Innovation », dans D'Agostino, G., A. Gaon et C. Piovesan (réd.), *Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law*, Toronto, ON, Thomson Reuters Canada.
- Scassa, T., 2021b. « Rights in Data, the Public Interest, and International Trade Law », dans Borchert, I. et L. A. Winters (réd.), *Addressing Impediments to Digital Trade*, Londres, Royaume-Uni, Centre for Economic Policy Research Press.

- Scassa, T., 2021c. « AI and Data Protection Law », dans Martin-Bariteau, F. et T. Scassa (éd.), *Artificial Intelligence and the Law in Canada*, Montréal, QC, LexisNexis.
- Schmidt, V., A. Luccioni, A. Lacoste, et T. Dandres, s.d. Machine Learning CO2 Impact Calculator. Adresse : <https://mlco2.github.io/impact/> (consulté en décembre 2021).
- Schwartz, O., 2019. « In 2016, Microsoft's Racist Chatbot Revealed the Danger of Online Conversation ». *IEEE Spectrum* (25 novembre).
- Schwartz, R., J. Dodge, N. A. Smith, et O. Etzioni, 2020. « Green AI », *Communications of the Association for Computing Machinery*, vol. 62, n°12, p. 54-63.
- Science International, 2015. *Open Data in a Big Data World*, Paris, France, International Council for Science (ICSU), International Social Science Council (ISSC), The World Academy of Sciences (TWAS), InterAcademy Partnership.
- Seglen, P. O., 1997. « Why the impact factor of journals should not be used for evaluating research », *British Medical Journal*, vol. 314, p. 498-513.
- Selbst, A. D. et S. Barocas, 2018. « The Intuitive Appeal of Explainable Machines », *Fordham Law Review*, vol. 87, p. 1085-1139.
- Sethi, A., A. Sankaran, N. Panwar, S. Khare, et S. Mani, 2018. *DLPaper2Code Auto-Generation of Code from Deep Learning Research Papers*, communication présentée dans le cadre du The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA.
- Simonite, T., 2018. « AI is the Future—But Where Are the Women? ». *WIRED* (17 août).
- Smith, D., 2021. « Here's What Died on the Order Paper ». *National Magazine* (17 août).
- Smith, M., 2017. « How Canada has emerged as a leader in artificial intelligence ». *University Affairs* (6 décembre).
- Somers, J., 2021. « The Pastry AI that Learned to Fight Cancer ». *The New Yorker* (18 mars).
- Sparkes, A., W. Aubrey, E. Byrne, A. Clare, M. N. Khan, M. Liakata, ... R. D. King, 2010. « Towards robot scientists for autonomous scientific discovery », *Automated Experimentation*, vol. 2, p. 1.
- Sparkes, M., 2021. « World's first 3D printed steel bridge opens in Amsterdam ». *New Scientist* (15 juillet).
- Stach, E., B. DeCost, A. G. Kusne, J. Hattrick-Simpers, K. A. Brown, K. G. Reyes, ... B. Maruyama, 2021. « Autonomous experimentation systems for materials development: A community perspective », *Matter*, vol. 4, n°9, p. 2702-2726.
- StatCan – Statistique Canada, 2016. *Profession – Système de classification des professions (CNP) 2016 (693A), plus haut certificat, diplôme ou grade (15), situation d'activité (3), âge (13A) et sexe (3) pour la population active âgée de 15 ans et plus dans les ménages privés du Canada, provinces et territoires, régions métropolitaines de recensement et agglomérations de recensement, Recensement de 2016 – Données-échantillon (25 %)*, Ottawa, ON, StatCan.

- StatCan – Statistique Canada, 2021a. *Enquête canadienne sur l'utilisation d'Internet, 2020 – juin 2021*, Ottawa, ON, StatCan.
- StatCan – Statistique Canada, 2021b. *Accès à Internet au Canada, 2020 – mai 2021*, Ottawa, ON, StatCan.
- Steffler, J., 2016. « The Indigenous Data Landscape in Canada: An Overview », *aboriginal policy studies*, vol. 5, n°2.
- Stevens, R., J. Nichols, et K. Yelick, 2020. *AI for Science*, Washington, DC, U.S. Department of Energy.
- Stoyanovich, J., S. Abiteboul, et G. Miklau., 2016. *Data, Responsibly: Fairness, Neutrality and Transparency in Data Analysis*, communication présentée dans le cadre du International Conference on Extending Database Technology, Bordeaux, France.
- Strubell, E., A. Ganesh, et A. McCallum, 2019. « Energy and policy considerations for deep learning in NLP », *arXiv*, n°1906.02243v1.
- Taddeo, M., A. Tsamados, J. Cowsls, et L. Floridi, 2021. « Artificial intelligence and the climate emergency: Opportunities, challenges, and recommendations », *One Earth*, vol. 4, n°6, p. 776-779.
- Taylor, J., 2021. « I'm sorry Dave I'm afraid I invented that: Australian court finds AI systems can be recognised under patent law ». *The Guardian* (30 juillet).
- The Artificial Inventor Project, 2021. Home. Adresse : <https://artificialinventor.com/> (consulté en septembre 2021).
- The Royal Society, 2017. *Machine Learning: The Power and Promise of Computers that Learn by Example*, Londres, Royaume-Uni, The Royal Society.
- The Royal Society, 2019. *The AI Revolution in Scientific Research*, Londres, Royaume-Uni, The Royal Society.
- The Turing Institute – The Alan Turing Institute, 2021. *A Roadmap for Developing AI Systems Capable of Making Nobel Quality Scientific Discoveries*, Londres, Royaume-Uni, The Turing Institute.
- Thompson, E., 2021. « U.S. Technology Company Clearview AI Violated Canadian Privacy Law: Report ». *CBC News* (3 février).
- Thompson, K. et L. Lucarini, 2021. *CPPA: An in-depth look at the de-identification provisions in Canada's proposed new privacy law*. Adresse : <http://www.dentondata.com/cppa-an-in-depth-look-at-the-de-identification-provisions-in-canadas-proposed-new-privacy-law/> (consulté en février 2022).

- TMS – The Minerals, Metals & Materials Society, 2017. *Building a Materials Data Infrastructure — Opening New Pathways to Discovery and Innovation in Science and Engineering*, Pittsburgh, PA, TMS.
- Tortoise Media, 2020. *The Global AI Index — Methodology*, Londres, Royaume-Uni, Tortoise Media.
- Tortoise Media, 2021. Global AI Index Rankings. Adresse : <https://www.tortoisemedia.com/intelligence/global-ai/> (consulté en mai 2021).
- U.S. NRC – U.S. National Research Council, 2014. *Big Data in Materials Research and Development: Summary of a Workshop*. Washington, DC, The National Academies Press.
- UNESCO – Organisation des Nations Unies pour l'éducation, la science et al culture, 2018. Canada First to Adopt Strategy for Artificial Intelligence. Adresse : http://www.unesco.org/new/en/media-services/single-view/news/canada_first_to_adopt_strategy_for_artificial_intelligence/ (consulté en mai 2021).
- UNESCO – Organisation des Nations Unies pour l'éducation, la science et al culture, 2020. *Outcome Document: First Draft of the Recommendation on the Ethics of Artificial Intelligence*, Paris, France, UNESCO.
- UofT – Université de Toronto, 2021a. Indigenous Data Sovereignty. Adresse : <https://guides.library.utoronto.ca/indigenoustudies/datasovereignty> (consulté en décembre 2021).
- UofT – Université de Toronto, 2021b. Pilot Program Embeds Ethics Into U of T Undergraduate Technology Courses. Adresse : <https://www.utoronto.ca/news/pilot-program-embeds-ethics-u-t-undergraduate-technology-courses> (consulté en août 2021).
- USIDSN – United States Indigenous Data Sovereignty Network, s.d. Join The US Indigenous Data Sovereignty Network. Adresse : <https://usindigenousdata.org/join> (consulté en octobre 2021).
- van der Aalst, W. M. P., M. Bichler, et A. Heinzl, 2017. « Responsible data science », *Business & Information Systems Engineering*, vol. 59, n°5, p. 311-313.
- Van Roy, V., 2019. *AI Watch National Strategies on Artificial Intelligence — A European Perspective in 2019*, Luxembourg, Luxembourg, Commission européenne.
- Vayena, E. et A. Blasimme, 2017. « Biomedical Big Data: New Models of Control Over Access, Use and Governance », *Journal of Bioethical Inquiry*, vol. 14, n°4, p. 501-513.
- Veale, M. et R. Binns, 2017. « Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data », *Big Data & Society*, vol. 4, n°2.
- Verma, I. M., 2014. « Editorial expression of concern: Experimental evidence of massive-scale emotional contagion through social networks », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, n°29, p. 10779.
- Villeneuve, S., G. Boskovic, et B. Barron, 2019. *Rebooting Regulation: Exploring the Future of AI Policy in Canada*, Toronto, ON, Canadian Institute for Advanced Research.

- Wainberg, M., D. Merico, A. Delong, et B. J. Frey, 2018. « Deep learning in biomedicine », *Nature Biotechnology*, vol. 36, n°9, p. 829-838.
- Wang, D. et A.-L. Barabási, 2021. *The Science of Science*. Cambridge, United Kingdom, Cambridge University Press.
- Wang, W. Y., B. Tang, D. Y. Lin, C. X. Zou, Y. Zhang, S. L. Shang, ... J. S. Li, 2020. « A brief review of data-driven ICME for intelligently discovering advanced structural metal materials: Insight into atomic and electronic building blocks », *Journal of Materials Research*, vol. 35, n°8, p. 872-889.
- Web Foundation – World Wide Web Foundation, 2018. *Open Data Barometer – Leaders Edition*, Washington, DC, World Wide Web Foundation.
- Webb, M., 2020. « The Impact of Artificial Intelligence on the Labor Market », *Social Science Research Network*.
- Webster, P., 2015. « CIHR modifies virtual peer review amidst complaints », *Canadian Medical Association Journal*, vol. 187, n°5, p. E151-E152.
- WEF – World Economic Forum, 2018. *How to Prevent Discriminatory Outcomes in Machine Learning*, WEF.
- Weis, J. W. et J. M. Jacobson, 2021. « Learning on knowledge graph dynamics provides an early warning of impactful research », *Nature Biotechnology*, vol. 39, p. 1300-1307.
- West, S. M., M. Whittaker, et K. Crawford, 2019. *Discriminating Systems: Gender, Race, and Power in AI*, AI NOW Institute.
- Whittlestone, J. et A. Ovadya, 2020. « The tension between openness and prudence in responsible AI research », *arXiv*, n°1910.01170v2.
- Whittlestone, J., R. Nyrupe, A. Alexandrova, K. Dihal, et S. Cave, 2019. *Ethical and Societal Implications Algorithms, Data, and Artificial Intelligence: A Roadmap for Research*, Londres, Royaume-Uni, Nuffield Foundation.
- Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, ... B. Mons, 2016. « The FAIR Guiding Principles for scientific data management and stewardship », *Scientific Data*, vol. 3, p. 160018.
- Wolchover, N., 2018. « Machine Learning's 'Amazing' Ability to Predict Chaos ». *Quanta Magazine* (18 avril).
- Woodgett, J., 2018. Peer Review and the Real Cost of Popularity (or How to Get 640 Comments on an Article). Adresse : <https://jwoodgett.medium.com/peer-review-and-the-real-cost-of-popularity-or-how-to-get-640-comments-on-an-article-d9522d4cd668> (consulté en août 2021).
- Woolston, C. et J. M. Perkel, 2020. « Sliced, diced and digested: Science ready in minutes », *Nature*, vol. 588, p. S136-S137.

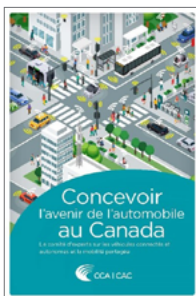
- WTO – World Trade Organization, 2017. *Agreement on Trade-Related Aspects of Intellectual Property Rights (as Amended on 23 January 2017)*, Genève, Suisse, WTO.
- Wu, B., T. Maxwell, et B. Crew, 2020. « Network effect: visualizing AI connections in the natural sciences », *Nature*, n°588, p. S112–S113.
- Wylie, B., 2018. *Open Data Endgame – Countering the Digital Consensus*, Waterloo, ON, Centre for International Governance Innovation.
- Xinhuanet, 2020. China Promotes Local AI Pilot Zones. Adresse : http://www.xinhuanet.com/english/2020-05/09/c_139043630.htm (consulté en septembre 2021).
- Xue, M., C. Yuan, H. Wu, Y. Zhang, et W. Liu, 2020. « Machine Learning Security: Threats, Countermeasures, and Evaluations », *IEEE Access*, vol. 8, p. 74720–74742.
- Yeston, J., 2021. Progress in Data and Code Deposition. Adresse : <https://blogs.sciencemag.org/editors-blog/2021/07/15/progress-in-data-and-code-deposition/> (consulté en juillet 2021).
- Ying, X., 2019. « An overview of overfitting and its solutions », *Journal of Physics: Conference Series*, vol. 1168, p. 022022.
- Young, E., J. Wajcman, et L. Sprejer, 2021. *Where Are the Women? Mapping the Gender Job Gap in AI*, The Alan Turing Institute.
- Yu, P. K., 2020. « The algorithmic divide and equality in the age of artificial intelligence », *Florida Law Review*, vol. 72.

Rapports utiles du CAC

Les rapports d'évaluation répertoriés ci-dessous sont accessibles sur le site Web du CAC (www.rapports-cac.ca):



Dynamiser la découverte
(2021)



Concevoir l'avenir de l'automobile au Canada
(2021)



De la recherche à la réalité
(2020)



Bâtir l'excellence
(2019)



Rivaliser dans une économie mondiale axée sur l'innovation : L'état de la R-D au Canada
(2018)



Paradoxe dissipé : Pourquoi le Canada est fort en recherche et faible en innovation
(2013)



Renforcer la capacité de recherche du Canada : La dimension de genre
(2012)



Honnêteté, responsabilité et confiance : Promouvoir l'intégrité en recherche au Canada
(2010)

Conseil d'administration du CAC*

Sioban Nelson, inf. aut., MACSS (présidente par intérim), professeure, Faculté des sciences infirmières, Université de Toronto; présidente, Académie canadienne des sciences de la santé (Toronto, Ont.)

Soheil Asgarpour, FACG, président, Petroleum Technology Alliance Canada; président désigné, Académie canadienne du génie (Calgary, Alb.)

Yves Beauchamp, C.M., FACG, vice-principal, administration et finances, Université McGill; président, Académie canadienne du génie (Montréal, Qc)

Chantal Guay, FACG, directrice générale, Conseil canadien des normes (Ottawa, Ont.)

Jawahar (Jay) Kalra, M.D., MACSS, professeur, Département de pathologie et de médecine de laboratoire et membre du Conseil des gouverneurs, Université de la Saskatchewan (Saskatoon, Sask.)

Cynthia E. Milton, vice-présidente associée à la recherche, Université de Victoria (Victoria, C.-B.)

Sue Molloy, Présidente de Glas Ocean Electric et professeure auxiliaire à l'Université Dalhousie (Halifax, N.-É.)

Proton Rahman, M.D., MACSS, professeur-chercheur universitaire, Faculté de médecine, Université Memorial (St. John's, T.-N.-L.)

Donna Strickland, C.C., MSRC, FACG, professeure, Département de physique et d'astronomie, Université de Waterloo (Waterloo, Ont.)

Julia M. Wright, MSRC, professeure, Département d'anglais, et professeure-chercheuse universitaire, Université Dalhousie; présidente, Académie des arts, des lettres et des sciences humaines, Société royale du Canada (Halifax, N.-É.)

*En janvier 2022

Comité consultatif scientifique du CAC*

David Castle, (président), professeur, École d'administration publique et Gustavson School of Business; chercheur en résidence, Bureau du conseiller scientifique principal du premier ministre du Canada (Victoria, C.-B.)

Maydianne C. B. Andrade, professeure de sciences biologiques, Université de Toronto à Scarborough; présidente, Réseau canadien des scientifiques noirs, Toronto, Ont.)

Peter Backx, MSRC, professeur, Département de biologie; titulaire, Chaire de recherche du Canada en biologie cardiovasculaire, Université York (Toronto, Ont.)

Stephanie E. Chang, professeure, School of Community and Regional Planning and Institute for Resources, Environment and Sustainability, Université de la Colombie-Britannique (Vancouver, C.-B.)

Neena L. Chappell, C.M., MSRC, MACSS, professeure émérite, Institute on Aging and Lifelong Health et Département de sociologie, Université de Victoria (Victoria, C.-B.)

Jackie Dawson, titulaire de la Chaire de recherche du Canada sur l'environnement, la société et les politiques et professeure agrégée au Département de géographie, Université d'Ottawa (Ottawa, Ont.)

Colleen M. Flood, MSRC, MACSS, directrice, Centre de droit, politique et éthique de la santé; professeure, Faculté de droit (section droit commun), Université d'Ottawa (Ottawa, Ont.)

Digvir S. Jayas, O.C., MSRC, FACG, professeur éminent et vice-recteur à la recherche et aux relations internationales, Université du Manitoba (Winnipeg, Man.)

Malcolm King, MACSS, directeur scientifique, Saskatchewan Centre for Patient-Oriented Research, Université de la Saskatchewan (Saskatoon, Sask.)

Chris MacDonald, professeur agrégé; directeur, Ted Rogers Leadership Centre; président, Département de droit et des affaires; Ted Rogers School of Management, Université Ryerson (Toronto, Ont.)

Barbara Neis, C.M., MSRC, professeure distinguée, titulaire de la chaire John Paton Lewis, Université Memorial de Terre-Neuve (St. John, T.-N.-L.)

Nicole A. Poirier, FACG, présidente, KoanTeknico Solutions Inc. (Beaconsfield, Qc)

Jamie Snook, directeur général, Torngat Wildlife Plants and Fisheries Secretariat (Happy Valley-Goose Bay, T.-N.-L.)

David A. Wolfe, professeur de sciences politiques, Université de Toronto à Mississauga; codirecteur, Innovation Policy Lab à la Munk School of Global Affairs and Public Policy, Université de Toronto (Toronto, Ont.)

Nous sommes sincèrement attristés par la perte de **Jeffrey A. Hutchings, MSRC**, titulaire de la chaire Izaak Walton Killam Memorial et professeur de biologie à l'Université Dalhousie. Il était aussi un membre important du comité consultatif scientifique. Nous adressons nos plus sincères condoléances à sa famille, ses amis et ses collègues.

*En janvier 2022



Council of
Canadian
Academies

Conseil des
académies
canadiennes

180, rue Elgin, bureau 1401
Ottawa (Ontario) K2P 2K3
Tél: 613 567-5000
www.rapports-cac.ca