# État de la situation sur les impacts sociétaux de l'intelligence artificielle et du numérique

2024

**AXES**

Santé durable

Éducation et capacitation

Industrie 4.0, travail et emploi

Arts, médias et diversité culturelle

Sobriété numérique et transition socio-écologique

Droit, cyberjustice et cybersécurité

Éthique, gouvernance et démocratie

obvia

## Crédits :

**Axe Santé durable**

**Sous la responsabilité de :** Martin Cousineau et Marie-Pierre Gagnon

**En collaboration avec :** Marianne Ozkan, Alexandre Castonguay, Aude Motulsky, et Maxime Sasseville

**Axe Éducation et capacitation**

**Sous la responsabilité de :** Chris Isaac Larnder et Nadia Naffi

**En collaboration avec :** Viviane Vallerand, Otilia Holgado, Normand Roy, Simon Parent, Nathalie Glais, Bruno Poellhuber, Ann-Louise Davidson, Valéry Psyché, Janvier Ngnoulaye, Christian Desîlets, Laurence Lachapelle-Bégin, Arnold Magdelaine, Elaine Mosconi, Julie Voisin et Didier Paquelin

**Axe Industrie 4.0, travail et emploi**

**Sous la responsabilité de :** Julie Garneau et Christian Lévesque

**En collaboration avec** : Isabelle Roberge-Maltais et Benoît Payette

**Axe Arts, médias et diversité culturelle**

**Sous la responsabilité de :** Colette Brin et Véronique Guèvremont

**En collaboration avec :** Charlotte Tessier et Johanne Chanca

**Remerciements :** Camila Moreira Cesar, François Giard, Isabelle Henrion-Dourcy, Jocelyne Kiss, Aaron Liu-Rosenbaum et Fatiha Sadat

**Axe Sobriété numérique et transition socio-écologique**

**Sous la responsabilité de :** Christophe Abrassart et Stéphane Roche

**En collaboration avec :** Florent Begue et Alexandre Gourret

**Axe Éthique, gouvernance et démocratie**

**Sous la responsabilité de :** Allison Marchildon et Bryn Williams-Jones

**En collaboration avec :** Antoine Boudreau LeBlanc et Gabrielle Verreault

**Axe Droit, cyberjustice et cybersécurité**

**Sous la responsabilité de :** Vincent Gautrais et Anne-Sophie Hulin

**En collaboration avec :** William Auclair

**Direction scientifique :** Lyse Langlois

**Soutien à la coordination et à la recherche :** Félix-Arnaud Morin Bertrand

Produit avec le soutien financier des Fonds de recherche du Québec

Québec
Fonds de recherche – Nature et technologies
Fonds de recherche – Santé
Fonds de recherche – Société et culture

## Remerciements :

Nous tenons à remercier le comité exécutif composé de Marc-Antoine Dilhac, Nathalie De Marcellis-Warin, Karine Gentelet, Philip Jackson, Steve Jacob, Lyse Langlois, Pierrich Plusquellec et Tania Saba pour leurs judicieux conseils.

DOI : 10.61737/CFRT3613
ISBN : 978-2-925138-46-4

Pour citer ce document : Obvia. (2024). État de la situation sur les impacts sociétaux de l'intelligence artificielle et du numérique - 2024. https://doi.org/10.61737/CFRT3613

# Table des matières


Introduction 4

Axe Santé durable 7

Axe Éducation et capacitation 13

Axe Industrie 4.0, travail et emploi 20

Axe Arts, médias et diversité culturelle 25

Axe Sobriété numérique et transition socio-écologique 32

Axe Droit, cyberjustice et cybersécurité 39

Axe Éthique, gouvernance et démocratie 46

Conclusion 52

Bibliographie 53

# Vue d'ensemble 2024 sur l'impact sociétal de l'IA

« Si l'on compare le monde moderne à celui du passé, la perte de l'expérience humaine est extraordinairement frappante. Ce n'est pas seulement, ni même principalement, la contemplation qui est devenue une expérience totalement dénuée de sens. La pensée elle-même, en devenant « calcul des conséquences », est devenue une fonction du cerveau, et logiquement on s'aperçoit que les machines électroniques remplissent cette fonction beaucoup mieux que nous. Le problème posé par les théories modernes du comportement n'est pas qu'elles soient fausses, mais qu'elles pourraient devenir vraies et qu'elles sont la meilleure conceptualisation possible de certaines tendances évidentes de la société moderne. »

(Hannah Arendt, *The Human Condition* (1958) : 321-322)

L'IA s'inscrit dans une révolution industrielle majeure et va transformer profondément la société et l'économie québécoises au cours des prochaines décennies. Cette technologie promet des avantages économiques et sociétaux importants. Cette transition numérique s'opère en même temps qu'une transformation environnementale, sociale et de gouvernance majeure et les nouveaux modèles qui en émergent sont disruptifs et changent le modèle économique et social.

Depuis 2019, le gouvernement québécois a consacré plus de 21,5 M $ à travers les Fonds de recherche du Québec pour l'étude des impacts sociétaux de l'IA. Ce concours scientifique a permis de mettre sur pied un Observatoire exclusivement dédié aux impacts sociétaux de l'IA et du numérique. Grâce à ce financement public, l'Obvia est devenue une voix indépendante et unique qui n'a pas d'équivalent dans le monde et qui suscite l'envie à l'international.

Dès sa création, cette nouvelle entité a su réunir plus de 250 professeurs provenant de 10 universités québécoises, 9 cégeps et collèges et plusieurs partenaires institutionnels dont des universités européennes, anglaises et coréennes. Ces partenariats ont été conclus par des ententes permettant aux chercheurs et aux étudiants gradués de collaborer avec leurs homologues dans des domaines comme l'éthique de l'IA, l'éducation, la cybersécurité, le droit, les cas d'usage et la mobilisation internationale. En plus de ces partenaires institutionnels, des initiatives de recherche ont mené à des collaborations avec des organisations internationales telles que l'OCDE, l'UNESCO, Interpol, l'Agence mondiale antidopage qui ont permis d'établir des partenariats de recherche tout en échangeant des expertises intéressantes.

La recherche menée au sein de l'Obvia prend appui dans les trois secteurs que sont les sciences humaines et sociales, les sciences de la santé ainsi que les sciences naturelles et le génie. Par le biais d'une démarche critique, les travaux qui y sont menés visent à identifier les enjeux sociétaux de l'IA et du numérique tout en contribuant à des solutions qui placent les êtres vivants et la biosphère au centre de leur cycle de développement et d'utilisation.

Cette mission guide les programmes et les initiatives de recherche mis de l'avant. L'orientation prise par l'Obvia au fil des années fait ressortir l'importance de prendre en considération dès la conception des systèmes d'IA, la contribution scientifique des sciences humaines et sociales, et sa capacité à mobiliser tous les acteurs pertinents à cette démarche.

Plusieurs discours circulent depuis plusieurs années à l'effet que l'IA doit être au bénéfice des sociétés ou de l'humanité, respectueuse de la diversité. Ces finalités exigent la circulation de concepts et méthodologies selon une approche interdisciplinaire. Pour que l' IA soit au service de la société, une collaboration étroite est nécessaire entre les gouvernements, les entreprises, le monde universitaire et la société civile. Il est nécessaire de mettre en place les bonnes réglementations, d'adopter des principes éthiques sains, de sensibiliser et d'éduquer le public, et de promouvoir une gouvernance transparente et inclusive. En combinant ces approches, il est possible de maximiser les avantages de l'IA tout en minimisant ses risques dans le but d'une contribution positive à la société.

La contribution des sciences humaines et sociales représente l'élément clé pour assurer un alignement en phase avec les besoins et les valeurs sociétales car un des enjeux actuels réside dans la complexité d'aligner les valeurs et les besoins sociétaux en phase avec la conception et le déploiement technologiques. L'apport des sciences humaines et sociales est souvent sous-estimé et les concepteurs technologiques, par des données probantes interdisciplinaires, sont invités à mettre toute leur expertise pour résoudre ce défi tout en faisant preuve d'humilité quant aux limites réelles de ces systèmes.

Les promesses et les investissements majeurs qui ont été faits par le gouvernement québécois depuis les dernières années ont été assez importants en IA. De 2017 à 2023, des contributions gouvernementales de près de 700 millions de dollars (Conseil de l'innovation du Québec, 2023) ont été versées à la recherche et au développement en matière d'intelligence artificielle (IA) ainsi qu'à son adoption. De ce montant, 402 millions de dollars ont été accordés pour favoriser le développement de technologies en appui à l'IA et : les activités de recherche; l'augmentation de la puissance de calcul; l'adoption de l'IA en entreprise; l'attraction des chercheurs en IA au Québec; l'élargissement de l'offre de formation dans ce domaine. Des partenaires clés au Québec se sont vu attribuer 164 millions afin d'accélérer l'intégration de l'IA dans les entreprises et d'appuyer les incubateurs et accélérateurs technologiques. Dans le cadre de la Stratégie québécoise de recherche et d'investissement en innovation, 125 millions ont été investis dans le soutien à l'excellence de la recherche universitaire en IA, la concrétisation de projets d'adoption de l'IA en entreprise, l'incubation et l'amorçage d'entreprises en IA et la prise en compte d'impacts sociétaux de l'IA. À ces montants s'ajoutent 500 millions de dollars que le gouvernement fédéral a injectés en IA au Québec depuis le milieu des années 2010 (PwC, 2021). Par ailleurs, un montant de 1,5 milliard de dollars en fonds de capital de risque en IA a été injecté par l'entremise de différents fonds d'investissement, selon une estimation de PwC (2021).

La scène nationale et la scène internationale voient apparaître plusieurs parties prenantes impliquées dans le développement et le déploiement en IA. L'Obvia demeure très à l'affût de ce qui s'y passe et plusieurs membres font partie de comités d'experts ou d'initiatives et jouent un rôle clé dans ces groupes. Cet état de la situation est donc teinté de ce qui a marqué la présente année. Au printemps 2023, le gouvernement du Québec a mandaté le Conseil de l'innovation du Québec (CIQ) afin d'initier une réflexion collective sur l'encadrement de l'IA au Québec. Les travaux du CIQ ont permis à de nombreux experts de s'exprimer. Parmi ceux-ci, plusieurs membres de l'Obvia ont été invités à mener des travaux en consultant plusieurs parties prenantes provenant d'organisations publiques, privées, du milieu académique et de la société civile. Sous la direction de l'innovateur en chef du Québec, un rapport (Conseil de l'innovation du Québec, 2024) a été déposé au ministère de l'Économie, de l'Innovation et de l'Énergie en février 2024 visant à offrir des recommandations pour la société québécoise. Presque en même temps, IVADO, un consortium de recherche, de formation et de mobilisation des connaissances en intelligence artificielle, porté par l'Université de Montréal, recevait en avril 2023 une importante subvention du Fonds d'excellence en recherche Apogée Canada d'une valeur de 124,5 M $. Ce financement a permis de lancer un vaste projet de recherche intitulé IAR3 dont les ambitions sont de bâtir l'avenir avec une IA robuste, raisonnante et responsable.

Sur le plan international, le Royaume-Uni organise le tout premier Sommet sur la sécurité de l'IA à Bletchey Park en novembre 2023 dans le but d'avoir une compréhension commune des risques que pose la technologie. Ce Sommet a rassemblé les gouvernements mondiaux et les puissants de l'industrie technologique venus discuter de la réglementation et réaffirmer l'importance de mieux encadrer les risques à la suite de la prolifération rapide des modèles d'IA notamment ceux derrière les systèmes d'IA générative comme ChatGPT.

Malgré une déclaration commune signée par 28 États dont l'Union européenne, les États-Unis et la Chine, ce rassemblement politique n'a pu dissiper un certain nuage qui traînait à la tête de ce Sommet, soit une discussion ouverte sur les failles et le lancement de modèles massifs, déployés par l'industrie technologique eu égard aux impacts sociétaux. À cet égard, un article de l'Institut de Relations internationales et stratégiques décrit de la manière suivante les orientations de l'industrie lors de ce sommet :

*Ainsi, les acteurs de la Big Tech s'affichent volontiers favorables à une volonté de réglementation qui se concentrerait sur les risques apocalyptiques pour l'humanité, venant des innovations de « frontière » et moins sur leurs propres modèles.* (Bourgeot, 2023)

Sur fond de possibles risques existentiels (*substantial risks*) pour l'intégrité de nos sociétés est ressortie de ce Sommet l'importance de garantir la sécurité tout en proposant des systèmes robustes. Sécurité et robustesse sont devenues les nouveaux créneaux repris par les gouvernements et les géants de la tech. Malgré certaines avancées, ce Sommet n'a pas permis de comprendre concrètement le rôle et les décisions des gouvernements en ce qui concerne les mesures qui seront prises pour atténuer ces risques ni les sanctions qui devront être appliquées envers les organisations qui font fi des bonnes pratiques. Quels sont ces risques substantiels, voire existentiels que posent les systèmes d'IA et quelles mesures quels risques seraient susceptibles de les atténuer par la mise en œuvre de mesures spécifiques? Ces questionnements avaient déjà fait l'objet de discussions au Québec sous l'impulsion de Yoshua Bengio lors d'un premier dialogue interdisciplinaire tenu en août 2023. En collaboration avec MILA, Obvia et IVADO, un panel a été réuni visant à engager une réflexion collective et interdisciplinaire sur les enjeux et risques posés par les récents développements de l'IA (Bengio et al., 2024). Cette initiative faisait suite à la sortie d'une lettre par l'institut Futur of Life réclamant une pause de six mois des recherches sur l'IA le temps de reprendre le contrôle de ces outils (Future of Life Institute, 2023). Cette initiative repose sur une conception du risque inscrite dans un futur incertain ayant le potentiel de dégrader le bien-être de la majeure partie de l'humanité, par exemple en détruisant la civilisation moderne. Lors de ce débat, Jocelyn Maclure nous invitait à la prudence en insistant sur l'importance de distinguer cette notion de risque posé par les systèmes d'IA tels qu'ils existent à l'heure actuelle avant de se lancer dans des scénarios catastrophes (Maclure dans Bengio et al., 2024). Ces risques existentiels ont été souvent propulsés par le courant long-termiste[1] qui exerce une influence certaine au sein de la communauté de l'IA.

Cette transformation qu'opère la technologie sur nos vies et nos habitudes s'insinue graduellement, doucement sans débat de société et représente à notre avis un risque bien réel. Un peu comme un spectateur, on assiste autant à une perte lente de capacité, autant à des gains de compétence. Dans certains cas, on accepte implicitement de suspendre certaines capacités. L'apprentissage machine (*machine learning*) remplace même certaines pratiques professionnelles parfois pour le mieux, parfois pour le pire. Les systèmes d'IA pénètrent les institutions publiques et privées et exercent une influence sur les processus décisionnels dont les répercussions sur nos vies sont encore à découvrir. Avec l'arrivée des technologies de traitement automatique des langues (*natural language processing* ou NLP), de même que l'analyse d'image ou de vidéo qui révolutionnent nos modes de vie, on met en exergue l'importance de s'attarder plus finement aux impacts sociétaux.

(...) un des enjeux actuels réside dans la complexité d'aligner les valeurs et les besoins sociétaux en phase avec la conception et le déploiement technologiques.

Les limites de ces systèmes sont peu discutées et sont souvent lancées sans être complètement optimales sur le plan de la sécurité , de la robustesse et des risques qui peuvent advenir. Un exemple de conséquences sous-estimées est les comportements toxiques sur les réseaux sociaux. Comme l'affirmait en 2017 Richard Khoury : « Pour l'instant, il n'y a pas vraiment d'outil pour surveiller et faire respecter les lois et notre sécurité sur Internet. » (Khoury, 2017). Sept ans plus tard, ce constat est toujours d'actualité, et le risque accru par l'exploitation des données massives.

Ainsi l'objectif de cet état de la situation 2024 sur les impacts sociétaux de l'IA et du numérique consiste à faire le point sur les connaissances actuelles en fonction des recherches poursuivies au sein de la communauté scientifique de l'Obvia et à présenter un point de vue équilibré et critique qui découle des travaux interdisciplinaires qui y sont menés. L'intention de ce premier état de la situation de l'Obvia est de mettre en évidence les avancées et les lacunes actuelles, tout en offrant des orientations qui prennent en compte les impacts sociétaux de l'IA et du numérique. Sans être une revue de la littérature exhaustive, cette vue d'ensemble se veut surtout un portrait des principaux enjeux soulevés par les travaux menés par l'ensemble de la communauté scientifique ainsi que par les axes de recherche de l'Obvia sous le leadership des coresponsables. Il a l'ambition de soulever des questionnements scientifiques sur la place qu'occupe présentement la technologie dans les différentes sphères de nos vies tant personnelles que professionnelles.

Au nom des membres du comité exécutif, je tiens à remercier les coresponsables pour avoir coordonné la production de cet état de la situation. Je suis très reconnaissante et les remercie sincèrement d'avoir participé avec enthousiasme à cette production scientifique qui sera mise à jour et produite chaque année.

**Lyse Langlois**
Directrice générale de l'Obvia

1 Voir entre autres *What We Owe the Future* de William MacAskill (2022)

# Axe
# Santé durable

**Sous la responsabilité de :**
Martin Cousineau et Marie-Pierre Gagnon

**En collaboration avec :**
Marianne Ozkan; Alexandre Castonguay; Aude Motulsky et Maxime Sasseville

# Actualité scientifique

## Des conditions favorables à l'IA dans un contexte règlementaire en transformation

La pandémie de Covid-19 a cruellement mis en lumière la vétusté du réseau de santé numérique au Québec. Le réseau comprend environ 10,000 systèmes d'information distincts (Québec, 2022). La circulation des données s'avère difficile, entravant la prise de décision basée sur des données probantes en temps opportun. Pour transformer ce système qui « date d'un autre siècle » comme le qualifie le gouvernement du Québec, ce dernier a lancé en 2022 un vaste chantier de modernisation numérique où la mobilité des données prend une place centrale (Québec, 2022). Le manque d'infrastructures adéquates et les difficultés d'accès aux données de qualité ralentissent l'adoption de l'IA en santé au Québec, mais aussi ailleurs (Hughes, 2023; Benaich, 2023; SciencesPo-Chaire Santé, 2023). La pénurie de personnel de soins et de TI ainsi que le décalage des programmes de formation contribuent à exacerber ces difficultés (Roppelt et al., 2024). L'incertitude du contexte réglementaire y contribue également (Roppelt et al., 2024). L'indice de maturité en IA, lequel s'avère au plus bas au Canada selon une étude, permettra de cibler des actions précises afin de rehausser les conditions favorables à l'adoption de l'IA (Castonguay et al., 2024).

Au Canada, le projet de loi C-27 incluant la *Loi sur l'intelligence artificielle et les données*, de même que les exigences du processus d'homologation de dispositifs médicaux basés sur l'apprentissage automatique sont en cours d'élaboration (Santé Canada, 2023). Au Québec, la *loi 5 sur les renseignements de santé et services sociaux et modifiant diverses dispositions législatives* a été adoptée en 2023. Elle vise notamment l'optimisation de l'utilisation des renseignements de santé. Elle définit les règles de gouvernance des renseignements de santé et initie la mise en place d'un système national de dépôt des renseignements. On peut penser que la loi aura des impacts, notamment sur l'accès aux données pour des fins de recherche. Les difficultés liées à l'établissement d'un régime de responsabilité en cas de dommages ou préjudices causés lors de l'utilisation d'un système d'IA par un professionnel de la santé ajoutent à l'incertitude réglementaire globale (OCDE, 2024). Avec l'entrée en scène des modèles de fondation, la question de la responsabilité devient d'autant plus complexe à cause de l'ajout de développeurs d'applications en aval des fournisseurs de modèles de fondation (Bommasani et al., 2022). L'Organisation mondiale de la santé propose des pistes à considérer en vue de responsabiliser les différents acteurs en jeu (OMS, 2024).

## Les modèles d'IA

L'IA à usage général[2] a dominé la scène dans le secteur de la santé en 2023 (Global Digital Health Monitor, 2023). L'IA à usage général est envisagée pour accélérer et améliorer tant les processus de soins que l'administration des soins. Elle est versatile, permet un accès rapide aux données, tant cliniques que médicales et offre une capacité à interagir avec l'utilisateur.

Les grands modèles de langage (GML, en anglais *large language models*) peuvent accélérer les tâches administratives telles que l'annotation et la mise à jour du dossier patient. Ils peuvent également aider les professionnels à établir un diagnostic ou à élaborer un traitement personnalisé et arrimé aux dernières avancées médicales et pharmaceutiques (Zhang et Boulos, 2023). Les GML sont également envisagés en soutien au patient. Les agents conversationnels fondés sur les GML peuvent répondre aux questions des patients, traduire et vulgariser le jargon médical. Ils pourraient être utilisés dans une démarche d'autosoin dans la gestion des maladies chroniques, pour atténuer certains troubles mentaux ou encore briser l'isolement (Ma et al., 2023; Balcombe, 2023). La promotion de saines habitudes de vie (activité physique, alimentation, sommeil, etc.) peut également reposer sur des assistants virtuels. En matière d'essais cliniques, les GML peuvent faciliter l'identification des patients en arrimant leur état de santé, historique ou traitement avec les objectifs de recherche (OMS, 2024 citant Medium, 2022).

Les avancées des grands modèles de vision (en anglais, *large vision models*) promettent d'améliorer les performances en imagerie médicale (Qiu et al., 2023). Les SAM (*Segment Anything Model*) permettent de discerner les tissus ou les structures anatomiques de façon automatique et suscitent un certain intérêt scientifique et clinique (Zahlan, 2023). La radiologie est un terrain particulièrement fertile pour l'IA. Sur les 691 dispositifs médicaux basés sur l'IA approuvés par la FDA en date d'octobre 2023, 77 % concernent la radiologie (Joshi, 2024). En matière de chirurgie, les grands modèles de vision peuvent être utilisés pour améliorer la précision de robots chirurgicaux. Les modèles d'IA générative sont envisagés pour produire et simuler des interventions chirurgicales en vue de perfectionner les habiletés des chirurgiens (Qiu et al., 2023).

2 Dans ce qui suit, l'IA à usage général est utilisée de façon interchangeable avec « modèles de fondation » (Marchildon et al., 2024). Les modèles de fondations peuvent être des grands modèles de langage, des grands modèles de vision ou des modèles multimodaux.

Finalement, dans le domaine de la biologie moléculaire, les performances des récents modèles d'IA utilisés pour prédire la structure, les fonctions, les interactions et les mutations de protéines pourraient accélérer considérablement la mise au point de médicaments et de vaccins (Qiu et al., 2023).

## Les risques des modèles d'IA à usage général

Le domaine de la santé représente un vaste terrain d'exploration pour les utilisations variées de l'IA à usage général (OMS, 2024). Plusieurs initiatives visant la mitigation des risques associés aux modèles de fondations ont vu le jour. L'OMS a consacré un rapport aux risques des modèles multimodaux (OMS, 2024). Le rapport identifie les risques par famille d'applications en santé et suggère des pistes d'actions pour les atténuer, selon les parties prenantes en cause. Les agents conversationnels peuvent entraîner notamment des risques de manipulation car leur faculté de personnaliser les conversations les rend persuasifs, ce qui pourrait être utilisé à mauvais escient (OMS, 2024). Les risques d'inexactitudes, d'erreurs et d'omissions causés par ce que l'on appelle couramment des « hallucinations » sont aussi documentés (OMS, 2024). Le flou juridique entourant la qualification des données produites par les agents conversationnels peut ajouter à l'incertitude quant à la confidentialité des données (Fournier-Tombs et McHardy, 2023). En Europe, le *Règlement sur l'intelligence artificielle* comprend des obligations spécifiques pour mitiger les risques de modèles à usage général (Parlement européen, 2024). Au Canada, les amendements au projet de loi C-27 prévoient des exigences spécifiques aux systèmes à usage général (Comité permanent de l'industrie et de la technologie, 2023)[3].

L'IA à usage général est envisagée pour accélérer et améliorer tant les processus de soins que l'administration des soins. Elle est versatile, permet un accès rapide aux données, tant cliniques que médicales et offre une capacité à interagir avec l'utilisateur.

## L'explication des résultats

L'explication algorithmique est la description du processus par lequel les données en entrée ont engendré les résultats obtenus en sortie (Doshi-Velez et al, 2019). L'explication renvoie à notre capacité à comprendre les résultats des algorithmes. Elle est étroitement liée à la transparence de l'algorithme et à la confiance que l'on accorde à sa fiabilité (Pégny et Ibnouhsein, 2018; Ribeiro et al, 2016). Or, certains modèles résistent à l'intelligibilité à cause de leur complexité liée notamment au nombre de variables incluses dans le modèle (Selbst et Barocas, 2018). Ces modèles complexes pourront exiger des explications approximatives. On parlera alors plutôt d'interprétation du modèle (Gilpin, 2019). Les modèles de fondation sont caractérisés par un volume sans précédent de données, de même que par une variété de formats (texte, image, vidéo) et une diversité de tâches, ce qui pose des défis d'interprétabilité additionnels par rapport aux autres modèles d'apprentissage (Bommasani et al., 2022).

## L'accessibilité et la qualité des données

L'accessibilité à des données de qualité s'avère un défi directement relié à la capacité de développer l'IA en santé (Panch, 2019). L'accessibilité des données requiert une infrastructure d'échange sécuritaire entre systèmes interopérables. La qualité des données est une condition essentielle à des systèmes fiables et à la réduction de biais statistiques.

Deux tendances ajoutent à la complexité du défi. La première est l'utilisation de l'IA à usage général. La conception et la mise à jour des modèles de fondation requièrent un volume de données auquel peu d'entités ont accès (Kak, 2023). Les coûts de curation des données et de la puissance de calcul requise freinent autant la conception que la mise à jour de modèles (Bender et al., 2021). Le modèle est parfois en décalage avec la réalité et peut perpétuer des stéréotypes (Bender et al., 2021). La deuxième tendance a trait à la démultiplication des sources de données. La variété des déterminants de la santé dans une vision globale de la santé (hygiène de vie, facteurs environnementaux) de même que les dispositifs médicaux connectés (par exemple, la télémétrie cardiaque pour une surveillance médicale accrue) ajoutent aux défis d'interopérabilité entre systèmes (Rani et al., 2023; Zahlan, 2023).

3 Voir la section de l'axe Droit, cyberjustice et cybersécurité pour davantage de détails à propos des différentes initiatives juridiques visant l'encadrement de l'IA.

# Débats scientifiques

## La protection de la vie privée

La démultiplication des flux de données augmente les risques de bris de confidentialité et d'atteinte à la vie privée (UNESCO, 2017). Les données, parfois sensibles, utilisées dans la conception et l'amélioration des modèles de fondations ne sont pas toujours transparentes et la protection dont elles font l'objet n'est pas toujours claire (Moodley et Rennie, 2023). Sachant que le consentement individuel spécifique perd en efficacité dans un contexte de données massives, diverses approches sont proposées en réponse au double objectif de partage des données et de protection de la vie privée (Barocas et Nissenbaum, 2014; Froomkin, 2019). Celles-ci consistent à revoir la portée du consentement ou l'exercice de consentement (UNESCO 2017; Cumyn, 2020; Ploug et Holm, 2016)[4]. Une autre approche consiste à déléguer le consentement à un fiduciaire dans une vision mettant de l'avant la valeur collective des données (Hulin, 2022). Certaines avancées technologiques pourraient aider à contenir les échanges de données. L'apprentissage fédéré, par exemple, permet d'entraîner un modèle à partir de données locales sans nécessité de transfert de données (Rani et al., 2023). L'IA générative limite le recours aux données de patients en générant des données synthétiques qui pourront compléter des jeux de données d'entraînement et y intégrer des cas rares (Zhang et Boulos, 2023).

Au-delà de la performance technique, la valeur même de l'IA en contexte clinique est incertaine. La tendance au techno-solutionnisme entraîne une surestimation de la valeur de l'IA qui pourrait contribuer à l'adoption de l'IA sans évaluation préalable rigoureuse (OMS, 2024).

## La valeur de l'IA

La performance technique des algorithmes est difficile à évaluer car seul 15 % des études rendent public leur code source (Benaich, 2023). Ceci pourrait en partie s'expliquer par le souci du secret commercial. Au-delà de la performance technique, la valeur même de l'IA en contexte clinique est incertaine. La tendance au techno-solutionnisme entraîne une surestimation de la valeur de l'IA qui pourrait contribuer à l'adoption de l'IA sans évaluation préalable rigoureuse (OMS, 2024). Les modèles économiques de l'IA sont peu abordés par la communauté scientifique (Roppelt et al., 2024). Au Québec, malgré d'importants investissements dans l'industrie de l'IA, ses bénéfices sont encore peu démontrés (Lavoie-Moore, 2024). Toujours au Québec, les dynamiques économiques de développement de l'IA « s'accordent difficilement » avec l'amélioration de la santé, selon le rapport de l'IRIS (Lavoie-Moore, 2024).

En radiologie, un domaine où pourtant la pénétration de l'IA est des plus importantes, il manque de données probantes sur la valeur de l'IA en contexte clinique (van Leeuwen et al., 2022). Toujours en radiologie, les modèles d'évaluation holistiques des systèmes d'IA se font rares (Boverhof, 2024). Boverhof suggère un modèle à plusieurs échelles, qui considère la valeur d'un système d'IA pour le patient (ie. améliore-t-il la santé du patient?), pour l'établissement de soins et plus largement à l'échelle communautaire (Boverhof, 2024). Une étude de l'Institut national d'excellence en santé et services sociaux (INESSS) relève l'hétérogénéité des approches d'évaluation et la nécessité de creuser plus à fond la perspective des usagers (Brahimi et Legendre, 2023). Silcox suggère la prise en compte des spécificités des communautés de patients, du phénomène de dérive de données (*data drift*) et de la durée requise pour mesurer la valeur prédictive à long terme des systèmes (Silcox et al., 2024).

4 On pourra penser au consentement global ou au méta-consentement, par exemple.

# Cas d'usage

## Woebot, l'agent conversationnel au service de la santé mentale

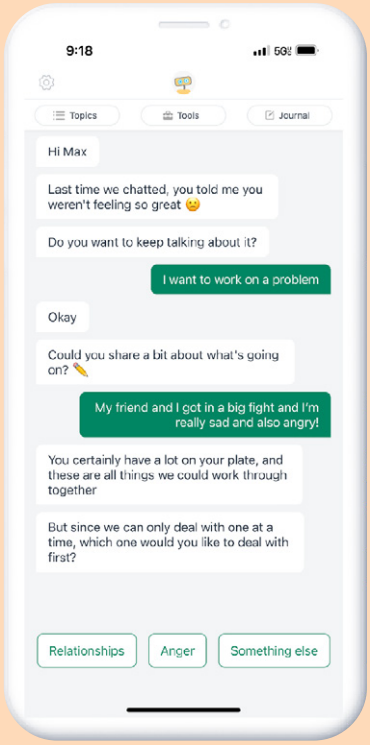


Photo : woebothealth.com

La pandémie de Covid-19 a eu des effets négatifs sur la santé mentale, notamment sur les enfants et adolescents (OMS, 2023). Cette situation, amplifiée par la pénurie de professionnels de soins, en incite plusieurs à se tourner vers l'IA en soutien à la santé mentale (Commission de la santé mentale du Canada, 2021). Utilisés comme complément thérapeutique ou encore en mode autonome, les agents conversationnels suscitent de l'intérêt (Commission de la santé mentale du Canada, 2021). Woebot, un agent conversationnel fondé sur l'IA, offre du soutien pour des troubles de dépression et d'anxiété (Woebot Health, s.d.). Disponible gratuitement sous forme d'application mobile, Woebot est recommandé par plusieurs établissements scolaires aux États-Unis (Keirerleber, 2022). Les agents conversationnels soulèvent par ailleurs de nombreuses inquiétudes, notamment en matière de sécurité des utilisateurs. La génération de réponses inadéquates par l'agent conversationnel peut entraîner de funestes conséquences en situation de crise (Vilaza et McCashin, 2021; Khawaja et Bélisle-Pipon, 2023). Or, les études sur l'efficacité des agents conversationnels sont peu nombreuses (Omarov et al., 2023; Balcombe 2023). Plusieurs observateurs réclament des études plus poussées en termes notamment de volumes des échantillons et d'impacts à long terme sur les utilisateurs (Omarov et al., 2023; Balcombe 2023). Une étude recense des failles méthodologiques au sein de nombreuses études sur l'utilisation de l'IA en santé mentale (Tornero-Costa et al., 2023). L'efficacité à long terme des applications en santé mentale est encore incertaine (Commission de la santé mentale du Canada, 2021).

# Pistes d'action et implications pour les parties prenantes

1

## Pour le législateur

**Favoriser la recherche sur** :

- l'IA générative afin (1)d'améliorer la compréhension de ses risques pour la santé, notamment les risques d'atteinte à la vie privée; (2)d'identifier les instruments existants favorisant la mitigation des risques (ex. lois, standards, directives); (3)d'identifier les zones d'ombre, les lacunes et recommander des ajustements à ces instruments.
- la valeur réelle de l'IA en contexte clinique : élaborer des modèles d'évaluation de l'IA, évaluer la valeur pour le patient, l'établissement et les communautés.

2

## Pour le ministère de la Santé et des services sociaux (MSSS)

**Poursuivre la mise en place des conditions nécessaires à l'adoption de l'IA**, notamment une infrastructure interopérable favorisant l'accès à des données de qualité.

3

## Pour les établissements de soins :

**Favoriser l'intégration de l'IA à risque faible**, par exemple les systèmes touchant les processus administratifs plutôt que les systèmes directement reliés aux soins.

Partant de la prémisse que les systèmes à faible risque soient soumis à moins de barrières à l'adoption que les systèmes à haut risque, ceci permettrait de livrer des bénéfices à court terme.

4

## Pour les ordres professionnels en santé :

**Poursuivre une veille sur les avancées de l'IA** pour outiller les professionnels en termes de formation ou lignes directrices (par exemple, sur l'utilisation d'agents conversationnels en santé mentale).

# Axe
# Éducation et capacitation

**Sous la responsabilité de :**
Chris Isaac Larnder et Nadia Naffi

**En collaboration avec :**
Viviane Vallerand, Otilia Holgado, Normand Roy, Simon Parent, Nathalie Glais, Bruno Poellhuber, Ann-Louise Davidson, Valéry Psyché, Janvier Ngnoulaye, Christian Desîlets, Laurence Lachapelle-Bégin, Arnold Magdelaine, Elaine Mosconi, Julie Voisin et Didier Paquelin

Les récents développements en intelligence artificielle ont propulsé les discussions sur ses applications en éducation et en formation au premier plan. Pour faire état des lieux, les membres de l'axe Éducation et capacitation ont choisi de mettre l'accent sur des enjeux actuels rassemblés autour de deux thématiques : le mariage entre la science des données et les sciences de l'apprentissage, et le potentiel capacitant des IA vues dans une approche sociotechnique critique (Collin et al., 2023) sous-jacente pour les personnes et les organisations. Le choix de ces deux orientations permet de mettre en relief, l'usage en éducation des données qui concernent des personnes apprenantes (taux de réussite, taux d'absentéisme, comportements d'apprentissage) et les impacts qui peuvent en découler, ainsi que les progrès de l'IA, notamment de l'IA générative (IAgen), qui posent de nombreuses questions sur la capacitation des actrices et acteurs éducatifs et des organisations en éducation.

# Actualité scientifique

## Sciences des données et sciences de l'apprentissage

### Transformation numérique du réseau scolaire et lois encadrant les données

Face à de nouvelles potentialités en termes de gestion efficace des données, diverses institutions scolaires vivent présentement de profondes transformations numériques dans le cadre du Plan Stratégique 2023-2027 mis en place par le ministère de l'Éducation. Cette stratégie vise à offrir plus efficacement des services auprès des élèves et étudiants du Québec, à faciliter la communication et à permettre au réseau scolaire de profiter de technologies émergentes (ministère de l'Éducation, 2023).

En focalisant toute son attention sur la transformation numérique de sa gestion des données, l'État a négligé son indispensable parallèle : la transformation numérique de la gestion de l'information gouvernementale (IG). Le rapprochement de ces deux gestions apparaît indispensable car c'est dans les documents que les données et l'IG prennent sens et que le discours démocratique de l'État s'incarne (Pacheco et al., 2023). Au-delà des énoncés de politiques et de l'acquisition des outils technologiques, la migration de la gestion documentaire (GD) analogique vers la gestion de l'information (GI) numérique piétine en l'absence de pratiques adaptées et affirmées, et de structures de suivi et de reddition de comptes.

Également, étant donné l'usage des données et des technologies numériques qui rendent possible la disponibilité des contenus traités par l'IA, il est pertinent de se questionner à savoir quelles sont les compétences numériques qui composent la littératie numérique des divers acteurs, préalables à l'intégration responsable de l'IA?

Or, les risques informationnels de la GI numérique, quand ils ne sont ni identifiés ni maîtrisés, s'invisibilisent et s'amplifient dans toutes les dimensions documentaires (Pacheco et al., 2023).

En parallèle, l'introduction de la *Loi modernisant des dispositions législatives en matière de protection des renseignements personnels* (communément appelée *Loi 25*) bouscule la transformation numérique et les technologies qui s'y rattachent comme les tableaux de bord. Les dispositions législatives de cette loi précisent les responsabilités des organisations en ce qui concerne la gestion des données personnelles ou de celles pouvant mener à l'identification d'une personne et porter atteinte à sa vie privée. Plusieurs questions se posent quant aux conséquences concrètes que cette loi peut avoir pour l'éducation et la formation : Quand et comment doit-on demander le consentement de personnes concernées lorsque leurs données personnelles sont utilisées par un système d'IA? Est-ce qu'une donnée générée par de l'IA à partir de données personnelles est considérée elle aussi comme « personnelle »? Les implications de cette législation suscitent encore plusieurs questions qu'il est nécessaire de débattre, il est possible que cette loi limite considérablement certaines utilisations de l'IA en éducation.

## Approche sociotechnique critique de l'IA pour les personnes et les organisations

### Les enjeux de l'IA pour le développement des personnes

La théorie de la capacitation de Sen (1999) explique ce qu'un individu peut réussir (ou pas), selon les ressources ou les opportunités qui sont à sa disposition. Dans le domaine de l'éducation, elle explique comment les personnes acquièrent des connaissances, des compétences et des capacités pour exercer un contrôle sur leur vie et réaliser leurs aspirations. Au prisme de cette théorie, l'avènement de l'IA amène à poser la question suivante : est-ce que l'IA est une opportunité ou un obstacle au développement continu des compétences des personnes?

Également, étant donné l'usage des données et des technologies numériques qui rendent possible la disponibilité des contenus traités par l'IA, il est pertinent de se questionner à savoir quelles sont les compétences numériques qui composent la littératie numérique des divers acteurs, préalables à l'intégration responsable de l'IA?

L'accessibilité du public à ChatGPT en novembre 2022 a bousculé le milieu de l'éducation (Davidson et Charbonneau, 2024) et semble avoir augmenté la polarisation face aux technologies déjà utilisées à des fins de formation. D'un côté, des parties prenantes socio-éducatives voient dans les applications de l'IA un nouveau moyen de permettre au personnel enseignant ou responsable de formation de profiler, de prédire, de tutorer, d'évaluer et de personnaliser de façon plus efficiente en se basant sur les besoins individuels des membres d'une communauté apprenante (Collin et Marceau, 2022). D'un autre côté, certains se préoccupent plutôt d'une augmentation des cas de plagiat, d'une amplification de la désinformation et du renforcement d'une paresse cognitive (CSE et CEST, 2024). Dans tous les cas, l'émergence de l'utilisation de l'IAgen, voire la conception de systèmes d'IA (SIA), aura des effets sur le milieu de l'éducation et de la formation qui devront être étudiés.

### Les enjeux de l'IA pour le développement des organisations

En parallèle, la formation à l'IA et son expérimentation à des fins technopédagogiques doivent être réalisées par et pour les actrices et acteurs éducatifs (personnes apprenantes, enseignantes, formatrices, administratrices, cheffes d'établissement, décideuses politiques ou encore du grand public). Elles sont nécessaires non seulement pour intégrer l'IA de façon responsable au réseau scolaire, mais aussi parce qu'elles permettent le développement de compétences de plus en plus prisées par les employeurs (Deschamps, 2024), tout en cherchant à former adéquatement les citoyen.e.s (Conseil de l'innovation du Québec, 2024). Il en va de même pour le besoin de former les travailleuses et les travailleurs à collaborer avec les outils d'IA[5], et ce, en vue d'une valorisation de la capacitation et d'une optimisation de la performance. Il s'agit d'ailleurs d'une des priorités gouvernementales, comme en témoigne un investissement de 81 millions $ pour la formation professionnelle dont celle à l'IA par le MEQ (Gouvernement du Québec, 2023a).

Il s'avère judicieux de définir les compétences à développer dans cette conjoncture, plusieurs actrices et acteurs appelant à une véritable initiative de littératie de l'IA et de mettre en avant des moyens adéquats pour y parvenir comme le souligne dans les rapports *Prêt pour l'IA* du Conseil de l'innovation du Québec (2024) et *Intelligence artificielle générative en enseignement supérieur : enjeux pédagogiques et éthiques* du Conseil supérieur de l'éducation (CSE) et de la Commission de l'éthique en science et en technologie (CSE et CEST, 2024). De plus, les écueils éthiques et critiques de l'IA, comme la reproduction des biais, les menaces à la vie privée et le manque de transparence algorithmique, semblent exacerbés avec l'arrivée de l'IAgen, qui produit aussi d'autres enjeux comme des hallucinations d'IA ou une dépendance cognitive (CSE et CEST, 2024). En pareilles circonstances, la recherche d'approches ou de stratégies de sensibilisation des parties prenantes à ces phénomènes s'impose de façon prioritaire, la recherche interdisciplinaire sera d'une grande aide à cet égard.

# Débats scientifiques

## Science des données et sciences de l'apprentissage

### Temps d'écran, hyperconnectivité et dépendance cognitive

Depuis plusieurs années que les débats de société se complexifient autour des enjeux du numérique, en tentant de considérer les tenants et aboutissants sur la santé physique ou psychologique, la société et l'éducation, etc. Ce faisant, les débats émergent en essayant d'avoir des consensus, et où nous retrouvons des amalgames des usages, des acteurs, des contextes et des enjeux, qui sont assez différents, tendant ainsi à proposer des interventions parfois trop simplistes, parfois trop complexes (Collins et al., 2024).

D'un côté, des chercheurs insistent sur les effets nocifs que les écrans ont sur les jeunes et sur la nécessité de limiter l'utilisation du numérique en général. Cette position semble être renforcée entre autres à la suite de la publication de la veille analytique « Écrans et hyperconnectivité » par l'Institut national en santé publique du Québec (INSPQ, 2024) et l'émergence de l'IAgen. Ici sont mis de l'avant les risques liés à un usage démesuré des technologies auprès des enfants. Cyberintimidation, perte d'habiletés cognitives, anxiété et même dépression, les effets négatifs du temps d'écran sont nombreux et bien documentés par le milieu scientifique (Fitzpatrick et al., 2023; INSPQ, 2024). Certains chercheurs critiquent vivement la place des technologies dans les écoles et valorisent plutôt un retour aux méthodes « papier-crayon ».

5 Voir la section de l'axe Industrie 4.0, travail et emploi.

C'est d'ailleurs de telles orientations que certains pays commencent de plus en plus à adopter comme la Suède qui suggère maintenant de limiter l'usage des technologies dans les écoles alors que par le passé, les plus jeunes étaient déjà exposés aux écrans (Planchon, 2023). Elle était reconnue comme un leader dans l'intégration des technologies dans le réseau scolaire (Courrier international, 2023) et prône dorénavant un usage plus tempéré. En France, l'utilisation des téléphones intelligents est interdite depuis plusieurs années (Dion-Viens, 2024) et aux États-Unis le débat continue de s'intensifier (Haidt, 2023). Plus près de chez nous, en Ontario et au Québec, les gouvernements semblent préconiser une telle approche en retournant aux éléments fondamentaux dans les salles de classe en Ontario (Ontario, 2024) ou en interdisant les téléphones intelligents dans les écoles primaires et secondaires du Québec (La Presse canadienne, 2023). Avec l'arrivée de l'IAgen, les personnes étudiantes et enseignantes du postsecondaire du Québec, craignent que cette nouvelle technologie n'amène la perte de capacités (Marquis et al., 2024), voire ne vienne accentuer la dépendance cognitive, c'est-à-dire à dépendre de l'IA et à perdre la capacité à réfléchir par soi-même.

D'un autre côté, certains chercheurs recommandent plutôt une utilisation des technologies plus équilibrée (Giroux et al., 2024), où les technologies incluant l'IA pourront dorénavant soutenir l'élève dans ses apprentissages par le biais de rétroactions courtes et directes, par exemple, ou encore soutenir le personnel scolaire pour mieux soutenir les élèves ayant des défis particuliers. Cette perspective s'aligne avec la théorie de capacitation de Sen (1999) comme quoi la réussite d'un individu dépend notamment de sa capacité à utiliser de façon responsable le numérique. Selon les chercheurs qui prônent cette approche, l'intégration responsable des technologies permettrait d'améliorer la qualité et l'accessibilité de l'éducation tout en soutenant la compétence numérique des jeunes Québécois. Par ailleurs, le ministère de l'Éducation et le ministère de l'Enseignement supérieur du Québec se sont dotés depuis 2019, d'un cadre de référence de la compétence numérique qui appelle les personnes enseignantes et étudiantes à développer leur littératie numérique (Gouvernement du Québec, 2019). Cela s'inscrit aussi en cohérence avec les besoins identifiés par le Conseil supérieur de l'Éducation et la CEST (2024) en matière d'intelligence artificielle. En ce sens, plusieurs recherches démontrent des effets positifs de l'utilisation d'outils numériques sur la motivation et certains aspects de l'apprentissage (Yu, 2021; Coban et al., 2022), notamment en écriture et en sciences (Zheng et al., 2016), sans compter des effets positifs sur les pratiques pédagogiques des enseignants (Tamim et al., 2015). Dans tous les cas, une approche prudente et réfléchie reste de mise puisque la recherche sur les effets positifs et délétères à long terme de l'IA, dont l'IAgen, notamment sur la dépendance cognitive des apprenants, reste à être approfondie.

Il apparaît donc essentiel d'approcher ces enjeux complexes à partir de perspectives multiples et complémentaires, afin de prendre le recul nécessaire pour mieux circonscrire ces phénomènes récents et émergents.

## Fracture et éloignement numérique

Dans un même ordre d'idées, l'accélération récente des usages du numérique a accentué la fracture numérique, soit les inégalités liées à l'accès et l'usage du numérique au sein de la population. Le milieu de l'éducation n'est pas à l'abri du phénomène en ne rendant pas toujours accessible à tous des outils technologiques dont des outils utilisant de l'IA et en n'intégrant pas dans les programmes de formation l'usage de ces outils (Plantard et al., 2023). Il reste, qu'encore une fois, il n'existe pas encore de consensus scientifique quant aux orientations à prendre pour limiter la fracture numérique actuelle. Faut-il faciliter, encadrer ou limiter l'accès et l'usage de ces technologies pour permettre une réelle équité?

L'éloignement numérique dépasse les approches centrées uniquement sur le manque d'accès aux technologies (fracture), de compétences (littératie) en y associant le concept de capabilité numérique (nature capacitante ou non-capacitante des technologies numériques) est en réalité grandement conditionnée par les conditions d'existence et le niveau de capital numérique des individus. Les résultats de recherche sur les usages de l'IA au prisme de ces différentes dimensions contribuent à la construction de politiques d'inclusion numérique (Plantard et al., 2023).

## Dilemme lié à l'usage des données éducatives

L'utilisation des données éducatives dans le domaine de l'intelligence artificielle en éducation pose plusieurs enjeux éthiques entre le respect de la vie privée et l'usage des données pour l'avancement des connaissances scientifiques en éducation. D'une part, la protection des données personnelles exige l'adoption de protocoles stricts, assurant l'anonymat des individus et garantissant à chacun le droit de se retirer des expérimentations en tout temps (Collin et al., 2023). Ces mesures sont cruciales pour maintenir la confiance des participants et respecter leurs droits fondamentaux. D'autre part, l'accès étendu à des données détaillées sur les apprenants et les enseignants représente un moyen pour exploiter pleinement le potentiel de l'IA en éducation (Ouyang et al., 2023). Ces données pourraient nous aider à mieux comprendre les processus d'apprentissage, d'améliorer la performance éducative, d'améliorer la conception des cours et d'augmenter les taux de réussite scolaire. Ainsi, le défi réside dans la création d'un cadre éthique et responsable qui maximise les bénéfices de ces données tout en préservant l'intégrité et la dignité de chaque individu impliqué.

## Usage de l'IA et approche sociotechnique critique par les organisations et les apprenants

### Enjeux éthiques en lien avec les systèmes prédictifs et décisionnels

Plusieurs gouvernements et institutions se tournent vers l'utilisation des tableaux de bord utilisant des IA prédictives et décisionnelles pour améliorer la prévention du décrochage scolaire et le soutien de la réussite. Les tableaux de bord sont des outils pouvant utiliser de l'IA qui permettent de représenter des données et des analyses pour soutenir la prise de décision des acteurs et actrices en éducation. En éducation, ils sont principalement utilisés pour l'analyse de données des apprenants et, si développés dans les règles de l'art, fournissent des résultats robustes (Sghir et al., 2023). Or, les implications de telles technologies ne vont pas sans risques pour lesquels une vigilance est de mise. Parmi eux, se trouve le renforcement de possibles biais à l'égard d'apprenants historiquement discriminés ou encore sous-représentés par les jeux de données qui pourraient être sur-représentés comme à risque de décrocher (Collin et al., 2023). Actuellement, des mesures de débiaisage[6] sont parfois incluses par ces équipes conceptrices à ces tableaux de bord pour mitiger ce risque (Chen et al., 2023). Toutefois, certains chercheuses et chercheurs notent que malgré la présence de biais, il reste important que des personnes apprenantes soient ciblées par les tableaux de bord afin de recevoir les ressources et le soutien nécessaires car le milieu de la recherche s'accorde pour dire que les groupes apprenants appartenant à des groupes marginalisés et/ou défavorisés sur le plan socio-économique ont moins de chance de réussite actuellement (Sghir et al., 2023).

Le recours à une grande quantité de données par ces systèmes d'IA actuels ne permet pas toujours de comprendre le fonctionnement de ces tableaux de bord, un phénomène que plusieurs appellent le phénomène de la boîte noire (Richard, 2018). Ce manque de transparence et d'explicabilité de ces systèmes d'IA peut créer une perte de confiance des utilisateurs envers ces outils. Le recours à des méthodes d'IA utilisant peu de données ou encore des méthodes d'explicabilité peut prévenir ce genre de situation (Khosravi et al., 2022). De plus, des stratégies et des actions pour augmenter le niveau de littératie des données des individus dans les organisations semblent fondamentales à un déploiement et usage responsable des données (Langlais et al., 2023), des éléments fondamentaux pour l'IA. Les enjeux face aux données et aux décisions d'automatisation et utilisation de l'IA dans les organisations sont davantage plus critiques pour les gestionnaires et les individus en position de leadership, leurs niveaux de littératie de données et de littératie numérique demeurent assez limités ayant un impact sur l'adoption des technologies et de l'IA, l'usage et la valorisation des données dans la prise de décisions (Mosconi et al., 2024).

### Hypertrucage, littératie informationnelle et numérique

La prolifération des hypertrucages (ou *deepfakes,* en anglais), identifiée par le Forum économique mondial comme un risque majeur à court terme (World Economic Forum, 2024), soulève de sérieuses préoccupations associées à la désinformation et à la manipulation numérique, toutes deux alimentées et propulsées par l'IAgen. Bien que des initiatives réglementaires contre les hypertrucages aient été mises en place pour contrer ce phénomène, telles que le *Projet de loi C-63* (2024), les standards C2PA (Coalition for Content Provenance and Authenticity, s.d.) et les efforts internationaux comme la coalition des 20 géants de la technologie (Gautherie, 2024), il est impératif de les compléter par des stratégies technopédagogiques ciblées. En effet, le potentiel d'utilisation malveillante des hypertrucages à des fins de déstabilisation s'accroît à un rythme alarmant, défiant les capacités des solutions technologiques de détection actuelles. Il est donc crucial de former les citoyennes et les citoyens, particulièrement dès le plus jeune âge (Collard, 2024), afin qu'elles et ils soient dotés d'un esprit critique permettant de déceler la désinformation véhiculée par les hypertrucages, ce qui relève encore une fois d'une littératie de l'IA (Naffi et Beaulieu, 2023). Malgré la hausse du nombre d'initiatives ayant pour objectif le développement de la littératie numérique chez les jeunes, une mise à jour continuelle de l'offre de formation s'impose afin de renforcer la résilience face à ce fléau et d'encourager l'adoption d'une posture active plutôt que celle d'une observation passive (Naffi et al., 2023). Le *programme Culture et citoyenneté québécoise* (CCQ) en est un exemple (ministère de l'Éducation du Québec, 2024), tout comme l'approche finlandaise qui opte pour une éducation aux médias (ministère des Affaires étrangères de Finlande, s.d; Ministry of Education and Culture of Finland, s.d;) continue, tout au long de la vie. La recherche scientifique menée dans ce champ est dynamique et il est souhaitable qu'elle se poursuive, de façon à mettre en lumière les pratiques et les initiatives dont la mise en œuvre est efficiente et les retombées positives.

6 Pensons ici à la mise en place de comités d'éthique, à l'utilisation de tests de biais, ou encore à une plus grande transparence des algorithmes.

# Cas d'usage

## Personnalisation de l'apprentissage par l'IA

L'application B12, développée par la compagnie québécoise Apprentx[7], améliore l'efficacité de la formation en entreprise et en milieu éducatif grâce à l'intelligence artificielle (IA) et aux sciences cognitives. Elle propose des parcours d'apprentissage personnalisés en fonction des besoins individuels, utilisant des techniques de micro-apprentissage et de renforcement adaptatif pour contrer la courbe de l'oubli et faciliter la mise en pratique des nouvelles compétences. L'IA de B12 permet de créer des activités de formation courtes et engageantes, intégrées dans le flux de travail quotidien via des outils comme Microsoft Teams, permettant aux formateurs d'ajuster les programmes grâce à des données analytiques. Des entreprises et institutions comme Winpak et le Collège Saint-Anne témoignent des gains significatifs en agilité et en efficacité grâce à cette plateforme.

Un nombre grandissant de personnes enseignantes pourraient avoir recours à l'IA, particulièrement l'IAgen, pour les soutenir dans le processus évaluatif, souvent considéré comme l'une des tâches les plus exhaustives chez ces professionnels (Celik et al., 2022). Cependant, le recours à l'IA pour évaluer pose plusieurs préoccupations, notamment la dégradation de la relation entre l'enseignant et l'élève et le changement du rôle de l'enseignant, qui pourrait se voir délesté de cette responsabilité, adoptant alors un rôle d'assistant aux apprentissages plutôt que celui de passeur des savoirs (CSE et CEST, 2024). Plus de recherche reste nécessaire pour évaluer, d'un côté, les impacts de ces changements dans le processus évaluatif des enseignants; et d'un autre coté, pour évaluer l'efficacité de l'IA pour les rétroactions « juste-à-temps » et la personnalisation de l'apprentissage (Psyché et Ruer, 2019).

7 Pour en apprendre plus sur l'application B12 : https://apprentx.rocks/fr/b12

# Pistes d'action et implications pour les parties prenantes

**Les pistes d'actions et implications s'inscrivent en cohérence avec les enjeux présentés, mais dépassent leurs frontières. Elles peuvent se regrouper en trois types : des ressources documentaires permettant de soutenir les acteurs et actrices du réseau de l'éducation; des formations liées aux besoins de développement de la compétence numérique à l'ère de l'IA; et l'implication concrète des personnes expertes dans les processus de mise en application de l'IA en éducation.**

1

## Des ressources ouvertes et libres à produire :

- Accélérer le transfert de connaissances multimodal via les données prometteuses et probantes issues de la la recherche interdisciplinaire de l'IA en éducation (p.ex. *Guides pratiques en technologies éducatives* publiés par EdTeq : Intégration[8] et Conception[9]).
- Impliquer les différentes parties prenantes (direction, personnes enseignantes, apprenantes, parents) dans le développement des politiques et des protocoles encadrant l'utilisation des systèmes d'IA et des données éducatives qui en résultent.
- Mettre en place des cadres réglementaires et éthiques adaptés encadrant l'utilisation de l'IA en éducation, en s'appuyant sur une veille des pratiques locales et internationales.
- Recueillir l'expression des besoins des parties prenantes puis cartographier les demandes et les attentes dans les différents programmes de formation (formation générale des jeunes, formation professionnelle, formation technique, enseignement supérieur, etc.) au regard du développement de la littératie de l'IA.

2

## Des dispositifs d'accompagnement à proposer :

- Mettre en place des programmes de formation continue des personnes enseignantes sur les concepts clés de l'IA, son fonctionnement et l'utilisation éthique et critique des technologies d'IA en classe.
- Sensibiliser les personnes apprenantes à l'IA, en les initiant à son fonctionnement et en développant leur esprit critique face à ces technologies.
- Détecter les initiatives existantes et les accompagner dans leur développement via la recherche participative.

3

## La présence de personnes expertes :

- Documenter les modalités concrètes mises en place avec des mécanismes de suivi et d'évaluation réguliers par les milieux d'enseignement ou l'entreprise pour l'implémentation d'une IA responsable.
- Concevoir des systèmes d'IA qui minimisent les biais algorithmiques, en diversifiant les jeux de données d'entraînement et en testant rigoureusement les performances sur différents groupes d'utilisateurs.
- Collaborer étroitement avec les établissements d'enseignement scolaires pour s'assurer de l'adéquation entre le projet d'établissement, l'organisation existante, la maturité de l'appropriation par les parties prenantes, les fonctionnalités des systèmes d'IA envisagés et les besoins pédagogiques révélés.
- Recommander l'investissement dans les systèmes d'IA explicables; et sur le développement technologique dans ce domaine qui pourrait faire émerger des systèmes qui évitent ou mitigent certaines difficultés de « sur-transparence ».

8 Voir Arseneau et al. (2024)
9 Voir Rioux et al. (2024)

# Axe

# Industrie 4.0, travail et emploi

**Sous la responsabilité de :**
Julie Garneau et Christian Lévesque

**En collaboration avec :**
Isabelle Roberge-Maltais et Benoît Payette

# Actualité scientifique

## L'IA et l'avenir du travail

La question des effets des changements technologiques sur le travail mobilise les chercheurs et chercheuses depuis de nombreuses décennies lors desquelles de multiples études et rapports de recherche ont montré les effets tantôt délétères des technologies sur les métiers et les tâches industrielles (Friedmann, 1956) tantôt plus subtiles par les transferts des compétences et des qualifications et de la répartition des temps de travail (Naville, 1961). Le chômage technologique illustrant le phénomène des pertes d'emploi pouvant en résulter est vite devenu un objet d'étude tant chez les économistes que les sociologues (Campa, 2017).

Aujourd'hui, l'introduction de l'intelligence artificielle (IA) comme outil de production dans les milieux de travail suscite à la fois de l'espoir et des craintes. Pourtant l'utilisation de l'IA dans les organisations ne semble pas être très répandue. Dans une étude récente, des chercheurs de l'Obvia montrent que si une assez forte proportion des entreprises au Québec utilise des technologies de pointe (42 %) ou émergentes (22 %), seulement 5 % d'entre elles utilisent l'IA (Dostie et Dufour, 2023). Cette faible pénétration de l'IA dans les entreprises est liée à différents facteurs, notamment à un faible investissement des entreprises en R-D (Conseil de l'innovation du Québec, 2024). Elle réside aussi sans aucun doute dans la difficulté à bien capter et mesurer le déploiement de l'IA. On a souvent tendance à définir l'IA comme un changement radical, mais elle se déploie de manière graduelle, par couche successive, souvent en arrière-scène, voire de manière invisible (Kellogg, et al., 2020). Cela pourrait certainement être l'une des explications pour laquelle environ un tiers des 3 000 professionnels du Québec interrogés par les chercheurs de l'Obvia révèlent ne pas être au courant de l'utilisation d'outils d'intelligence artificielle au sein de leurs entreprises ou organisations, bien que 86 % d'entre-eux affirment utiliser beaucoup les technologies numériques dans le cadre de leurs activités professionnelles (de Marcellis-Warin et Mondin, 2022).

En raison de la malléabilité et des trajectoires contrastées de l'adoption des nouvelles technologies selon les secteurs et les entreprises, il est extrêmement hasardeux de dégager des constats définitifs. L'impact des nouvelles technologies et de l'intelligence artificielle sur le monde du travail reste en grande partie à explorer. Cette incertitude souligne la nécessité d'une interprétation prudente des données de recherche actuelles et justifie la réalisation d'études détaillées pour saisir pleinement les répercussions des technologies naissantes sur le futur du travail.

# Débats scientifiques

## La fin du travail? Vraiment?

Le risque d'être remplacé par une IA est présenté dans divers médias comme un risque majeur pour le travail et l'emploi d'aujourd'hui et de demain. Les thèses de la suppression des emplois proviennent de certains rapports prospectifs qui sont pour plusieurs, alarmistes ou du moins, quelque peu excessifs puisque désincarnés des rapports réels de production socio-économiques. À titre d'exemple, pour certains, près de la moitié (47 %) des emplois aux É.-U. sont à risque (Frey et Osborne, 2013; Frey et Osborne, 2017) tandis que pour d'autres, tous les emplois sont à risque d'être remplacés (Ford, 2015). Ces études sont largement critiquées, car d'une part l'impact net est difficilement prédictible si l'on considère les nouveaux emplois qui seront créés par la mise en marché des nouvelles technologies et d'autre part, ce ne sont pas tant les emplois qui seront touchés, mais plutôt certaines de leurs tâches qui seront impactées. C'est à cette hypothèse que travaille d'ailleurs l'Organisation internationale du travail qui étudie l'impact de l'IA générative sur le travail via l'étude de la nature des changements plutôt que l'estimation des pertes d'emplois en mettant en lumière l'importance des contextes industriels et institutionnels, évitant ainsi les écueils d'une généralisation prospective (Gmyrek et al., 2023). Cela dit, bien qu'il devienne hasardeux de conclure sur l'impact réel de la perte d'emploi, ces différentes données nous éclairent sur le fait que l'ensemble des secteurs d'activités (fabrication, services, transport, tourisme, etc.) et des fonctions de l'entreprise (production, gestion de la chaîne d'approvisionnement, marketing, comptabilité, RH, contentieux, etc.) sont susceptibles d'être touchés par l'introduction des nouvelles technologies, et ce, à des degrés variables selon les secteurs et les institutions en place, ce qui se traduira vraisemblablement par une recomposition des professions et des métiers (Carbonell, 2022).

Si la thèse de la fin du travail semble plus difficile à mesurer, celle de la transformation du travail prend un espace considérable dans la littérature scientifique. En effet, en complément des études sur les pertes d'emploi, les regards sur les emplois restants font ressortir une forme de polarisation du marché du travail par le transfert des qualifications et la création de certains postes hautement qualifiés (ingénieurs, concepteurs), mais aussi de nombreux emplois moins qualifiés et moins valorisants (opérateurs, services) soumis à une machine ou une technologie (Autor, 2019; Autor et Dorn, 2013).

Ces observations amènent à un constat plus large sur la problématique du numérique et de l'IA sur le travail : la problématique ne concerne pas nécessairement la dimension quantitative du travail (nombre d'emplois), mais la dimension qualitative du travail.

## Oui, mais quels types de travail?

La qualité du travail devient donc névralgique dans le débat, car si l'objectif d'adoption de l'IA - comme de toute technologie - se concentre sur l'augmentation de la productivité (et de la rentabilité) alors l'humain en tirera-t-il une amélioration de ses conditions de travail? Cette question a suscité beaucoup d'intérêt au cours de la dernière décennie et les chercheurs ont proposé diverses typologies et cadres pour appréhender la qualité du travail dans le contexte du numérique (Berg et al., 2023; Gesualdi-Fecteau et al., 2023; Lévesque et al., 2024). En admettant que cette question est complexe et multidimensionnelle, il existe un consensus important sur la nécessité de prendre en compte un large éventail de facteurs économiques et non économiques pour évaluer les « bons et les mauvais emplois » (Guest et al., 2022; Warhurst et Knox, 2022). Pour résumer les différentes littératures, nous retenons les grandes catégories d'indicateurs qui tournent autour de trois dimensions :

- Risques au travail – ces risques peuvent être économiques, sociaux, physiques ou psychologiques.
- L'autonomie au travail selon que l'IA améliore ou réduit la capacité des travailleurs à prendre des décisions sur la séquence des tâches et les réalisations du travail.
- Le sens que les employés donnent à leur travail, s'il est significatif et épanouissant, et s'ils ont la possibilité de développer et de mobiliser leurs aptitudes et leurs compétences.

Selon les cas, ces dimensions sont complémentaires et quelquefois contradictoires. En effet, à titre d'exemple, bien que certaines nouvelles technologies permettent la réduction des risques physiques par le remplacement des tâches lourdes et répétitives, d'autres technologies accroissent le stress et intensifient les risques psychosociaux (stress, perte de motivation). Par ailleurs, l'autonomie accordée dans certains cas pour utiliser certaines technologies d'IA peut placer les employés devant une impossibilité à juger adéquatement de la véracité ou de la validité d'une décision de l'IA générative souvent opaque et inexplicable. Puisque l'articulation entre ces trois dimensions se déploie dans des espaces de travail où il existe différentes formes de gouvernance, il devient impératif de créer des espaces de gouvernance inclusive et participative pour discuter et diminuer les risques que posent ces technologies à la qualité du travail. Les résultats de recherche sont concluants à cet égard (Doellgast et Warner, 2022) : dans les pays, les secteurs et les entreprises où des processus et des régulations permettent aux employés et à leurs représentants d'intervenir à toutes les étapes du processus de mise en œuvre de l'IA, ses effets sur la qualité du travail sont plus positifs. Outre le milieu du travail, l'utilisation des comités paritaires ou d'autres formes de concertation sectorielle est une voie porteuse pour les acteurs patronaux et syndicaux (Dupuis, Murray et Wu, 2022). Pour l'acteur syndical, lorsqu'on compare les différentes stratégies, on voit combien ces stratégies de mobilisation diffèrent selon leur contexte historique et institutionnel, certains allant même jusqu'à participer activement dans des grappes industrielles et technologiques (Garneau et al, 2023). Tout ceci met en lumière les multitudes d'espaces qui s'offrent aux acteurs pour agir sur cette transformation. Pour le dire autrement, remettre les humains au cœur de l'IA c'est aussi prendre le parti de l'humain à la fois dans sa forme individuelle et collective.

La qualité du travail devient donc névralgique dans le débat, car si l'objectif d'adoption de l'IA - comme de toute technologie - se concentre sur l'augmentation de la productivité (et de la rentabilité) alors l'humain en tirera-t-il une amélioration de ses conditions de travail?

# Cas d'usage

## L'avenir du travail dans l'industrie aéronautique

L'industrie manufacturière de l'aérospatiale est vigoureuse et les gains en productivités sont cruciaux pour le maintien des différentes entreprises qui œuvrent dans cet environnement. Elle est à la pointe de l'innovation technologique, que ce soit en matière de produits, de matériaux ou de processus de production. L'introduction des nouvelles technologies de production rend possible le suivi de toutes les informations relatives à la fabrication en temps réel, par la captation des données provenant de robots, de moniteurs de machines et de rapports établis par les employés. Ces systèmes connectés permettent la collecte et le calcul d'une grande quantité d'informations sur le processus de production, y compris l'activité des travailleurs.

Deux scénarios pour le futur du travail dans l'aéronautique se dessinent alors. Dans le premier scénario, le travail des machinistes est enrichi par des tâches plus variées et plus complexes, leur conférant plus d'autonomie pour contrôler le rythme et la séquence du travail. Dans le second scénario, les tâches précédemment accomplies par le machiniste traditionnel, telles que la programmation et les ajustements, sont transférées au personnel des TI ou aux analystes de données, dont le travail s'en trouve enrichi, mais celui des machinistes et des opérateurs se trouve appauvri. Le travail du machiniste traditionnel se transforme en celui de surveillant de machine sous le nouveau titre d'opérateur. Dans ce second scénario, l'enjeu est moins quantitatif que qualitatif : si le machiniste conserve son emploi, son travail nécessite moins de compétences tacites basées sur une connaissance fine du processus de fabrication. En résulte une insatisfaction grandissante des machinistes à l'égard de leur travail qui se traduit par une augmentation du taux de roulement. La polarisation du marché interne de l'emploi provoquée par l'introduction des nouvelles technologies de production crée un dilemme pour les entreprises. Dans un contexte de pénurie de main-d'œuvre, la déqualification du métier de machiniste rend le travail moins attractif, ce qui aggrave encore plus le problème de pénurie.

Ces deux scénarios illustrent que l'avenir du travail n'est pas uniquement déterminé par la technologie. L'impact des nouvelles technologies sur le travail est socialement construit et peut mener à une recomposition ou une décomposition des métiers selon les circonstances sociales et les modalités de mise en œuvre des nouvelles technologies.

# Pistes d'action et implications pour les parties prenantes

**Dans un tel contexte, un des enjeux clés consiste à identifier les conditions susceptibles d'assurer que l'IA s'accompagne d'une amélioration qualitative des dimensions constitutives du travail. Pour les chercheurs de l'IA, le défi consiste à coconstruire avec les partenaires du marché du travail des recherches qui leur permettent de faire des choix éclairés. Ce défi peut se décliner sous la forme de trois questions de recherche :**

## 1

**Quelles sont les formes organisationnelles et institutionnelles susceptibles de favoriser une implication active d'une multitude de parties prenantes avec des expertises variées, (car les entreprises ne peuvent y arriver seules)? La tâche est colossale et plusieurs entreprises, notamment les PME n'ont pas les ressources internes pour y arriver.**

- À cet égard, il ne s'agit pas toujours de créer de nouveaux mécanismes, mais aussi d'utiliser ceux déjà à notre disposition. Par exemple, il faut bénéficier des expériences et réussites dans différents secteurs au Québec autour des comités sectoriels de main-d'œuvre pour réfléchir sur l'arrimage entre l'IA et le déploiement des nouvelles compétences. Ces mécanismes de concertation et de délibération unique en Amérique du Nord peuvent représenter une avenue importante pour favoriser le déploiement d'une IA inclusive, car ces comités disposent d'une expertise dans leur secteur respectif où les assemblées délibérantes et les acteurs patronaux et syndicaux sont légalement mandatés pour trouver ensemble des solutions.

## 2

**Quelles sont les ressources collectives et les capabilités requises pour que les parties prenantes disposent des outils pour agir en commun sur le déploiement de l'IA?**

- À cet égard, les États doivent mettre en place des programmes pour accompagner les entreprises et les travailleurs ainsi que des outils et des règles pour assurer la transparence et la protection des données qui sont colligées tout au long de la vie au travail (données de production et de gestion).

## 3

**Quelles compétences doivent être développées et mobilisées? Cette question est cruciale pour une mise en œuvre inclusive et responsable des nouvelles technologies et de l'IA dans les organisations.**

- Dans un premier temps, les acteurs des milieux de travail, incluant l'État, doivent réfléchir sérieusement à mettre en place des mesures pour favoriser le développement des compétences. En ce sens, la formation au cours de la durée de vie est un enjeu clé afin d'assurer une plus grande mobilité sur le marché par la formation continue dans les institutions d'enseignement et dans les entreprises.

- Dans un deuxième temps, la question des compétences soulève le défi de la perte potentielle des savoirs tacites lors de l'adoption de nouvelles technologies, augmentant ainsi le risque d'échecs techniques susceptibles de nuire à la réputation et à la confiance envers ces organisations.

Les impacts sociétaux du numérique et de l'IA sur le travail et l'emploi restent largement indéterminés et les acteurs du monde du travail sont en quête de données probantes et adaptées à leur contexte pour mieux comprendre comment l'IA façonne l'avenir du travail. Les chercheuses et les chercheurs de l'axe Industrie 4.0, travail et emploi veulent contribuer à cette tâche et accompagner les acteurs pour assurer une amélioration de la qualité du travail, mais aussi plus largement de la société.

# Axe

# Arts, médias et diversité culturelle

**Sous la responsabilité de :**
Colette Brin et Véronique Guèvremont

**En collaboration avec :**
Charlotte Tessier et Johanne Chanca

**Remerciements :**
Camila Moreira Cesar, François Giard, Isabelle Henrion-Dourcy, Jocelyne Kiss, Aaron Liu-Rosenbaum et Fatiha Sadat

# Actualité scientifique

## L'IA et la création de contenus artistiques et médiatiques

### Culture

Dans le domaine des industries culturelles et créatives, l'IA, et particulièrement l'IA générative, suscite des préoccupations majeures. Depuis que ces systèmes ont fait leur apparition dans la sphère publique dans le courant de l'année 2022, les secteurs de la musique, du cinéma, de l'édition et du livre, du jeu vidéo ou plus largement des arts visuels sont secoués tant par les bouleversements qu'ils engendrent que par ceux qui sont anticipés. L'IA générative a semé, et sème toujours, le doute dans le milieu de l'art (Baillargeon, 2022), à un point tel que les chercheuses et chercheurs sont amenés à s'interroger sur la notion même de créativité. À condition, bien entendu, de s'autoriser une association entre IA et créativité, d'où l'appellation d'« IA créatives[10] » utilisée par certains. Ainsi, une question hautement d'actualité est celle de savoir si l'IA peut faire preuve de créativité.

Les réflexions sur ce thème sont récentes, la créativité et l'IA étant initialement apparus comme des concepts étrangers l'un de l'autre; mais les avancées récentes en matière d'IA générative interrogent cette dissociation (Crimaldi et Leonelli, 2023). Des auteurs avancent que la notion d'IA créative est un contresens dans la mesure où les IA ne possèdent aucunement les principales caractéristiques humaines de la créativité, c'est-à-dire l'authenticité, l'originalité et l'émotivité. Certains parlent de créativité artificielle et ce, afin de la distinguer - ou à tout le moins de ne pas la mettre au même niveau - que la créativité humaine (Runco, 2023). D'autres mettent au contraire en avant la notion de co-créativité pour décrire un mélange de créativité humaine et artificielle (Wingström et al., 2024) ou encore de créativité humaine augmentée (Eapen et al., 2023).

Intimement liée aux questionnements qui précèdent est la réflexion sur la notion même d'auteur qu'engendre l'essor de l'IA générative dans le champ de la création artistique. Comme le résume Georges Azzaria : l'IA brouille la distinction classique entre la main et l'outil et nous amène à nous interroger sur l'interaction entre l'humain et la machine, notamment dans une perspective de droit d'auteur (Azzaria, 2023, p. 551). De conception anthropocentriste, le droit d'auteur canadien limite l'attribution du statut d'auteur aux personnes physiques[11] à l'origine de l'apport créatif dans l'œuvre créée (ou le résultat obtenu). En revanche, dans un cas comme celui du *Next Rembrandt* où le résultat n'était pas prévisible par les personnes ayant programmé les algorithmes, il est possible de soutenir que l'œuvre en question tomberait d'office dans le domaine public (Azzaria, 2023, p. 557).

L'IA générative génère de nombreuses autres questions, actuelles ou prospectives, en lien avec la place qu'elle pourrait progressivement se tailler dans le milieu des arts et de la culture. Par exemple : Assistera-t-on à une explosion de l'innovation assistée par l'IA? Les machines - et surtout ceux qui les contrôlent - pourraient-elles (ou pourraient-ils) en venir qu'à détenir le monopole de la créativité? Ou verra-t-on plutôt émerger un monde partagé entre, d'une part, ceux qui seraient enclins à payer pour de l'art authentique issu du génie créateur humain (*human-made premium*), par exemple identifié par un label (De Cremer et al., 2023) et, d'autre part, un monde composé d'œuvres créées par des IA? Et quels seront les impacts de ces changements sur la diversité de l'offre créatrice? Toutes ces questions restent encore en suspens et soulèvent de nombreux débats, non seulement sur la création et la définition même de l'art, mais aussi sur l'impact de cette révolution sur le statut de l'artiste et sur ses droits économiques, sociaux et culturels. Des questions se posent en outre sur la manière d'accompagner les artistes, mais aussi les industries culturelles et créatives, dans ce qui paraît être une inévitable adaptation aux changements qui ont déjà cours. Ultimement, c'est aussi le rôle de la norme qui est interrogé et sa capacité à encadrer le recours à ces technologies dans ce qui a, jusqu'ici, été qualifié de champ de la créativité humaine.

10 Pour en savoir plus sur le potentiel créatif des IA, voir Reynaud, O. (2018) L'intelligence artificielle créative...vers de nouveaux horizons, en ligne : https://olivierreynaud.medium.com/lintelligence-artificielle-cr%C3%A9ative-vers-de-nouveaux-horizons-efc588c761dd

11 Dans *Setana Sport Limited c. 2049630 Ontario Inc.*, la Cour Fédérale s'est prononcée sur la question de savoir si une personne morale pouvait être auteure au sens de la Loi sur le droit d'auteur : « Un auteur doit être un être humain ou un groupe d'êtres humains. Le droit d'auteur subsiste pendant la vie de l'auteur et dure encore cinquante (50) ans (voir la Loi sur le droit d'auteur, L.R.C. 1985, ch. C-42, article 6). La subsistance du droit d'auteur est prévue lorsqu'il y a plusieurs auteurs, où il est question de la vie du dernier survivant (article 9). Le concept de « vie » ne s'applique pas à une personne morale telle que Setana. Lorsque la Loi sur le droit d'auteur envisage la possibilité qu'une entité non humaine soit la partie à l'origine du droit d'auteur, elle le fait explicitement; c'est le cas notamment des radiodiffuseurs, des producteurs ou des propriétaires d'une épreuve négative d'une photographie. ». Pour consulter la décision : https://decisions.fct-cf.gc.ca/fc-cf/decisions/fr/item/54291/index.do

Des questions se posent en outre sur la manière d'accompagner les artistes, mais aussi les industries culturelles et créatives, dans ce qui paraît être une inévitable adaptation aux changements qui ont déjà cours.

Enfin, il est essentiel de souligner les préoccupations qui émergent en ce qui a trait à l'essor de l'IA dans le champ des industries culturelles et créatives et son impact environnemental[12]; la partie de ce rapport dédiée à l'axe Sobriété numérique et transition socio-écologique offre une vue d'ensemble des enjeux spécifiques reliés à ces enjeux.

## Médias

Pour le journalisme, les IA génératives soulèvent un questionnement fondamental sur l'exactitude de l'information et le rapport aux faits, voire à la réalité. Il y a quelques années à peine, le cadrage et le *retouchage* des photos de presse faisait encore l'objet d'un débat éthique (Koltermann, 2020) : mais demain, des personnages de synthèse pourront peut-être présenter un journal télévisé entier dont les reportages auront aussi été constitués par des systèmes d'IA à partir d'éléments (images, textes) glanés sur internet (grâce notamment au travail de « vrais » journalistes...)[13]. Le domaine de l'information, déjà menacé par le transfert des revenus publicitaires et la captation de l'attention par les grandes plateformes numériques, se dirige-t-il vers une crise existentielle encore plus profonde avec la progression fulgurante des outils d'intelligence artificielle?

Certains de ces outils sont déjà largement intégrés, parfois discrètement, dans les salles de presse et les médias, que ce soit en production (reconnaissance vocale et transcription, traduction, lecture vocale, aide à la synthèse et à la rédaction) ou en distribution (systèmes de recommandation plus ou moins personnalisés et optimisés pour les usagers) (Beckett et Yaseen, 2023). Le lancement de ChatGPT, lui-même très fortement médiatisé, à l'instar d'une couverture journalistique généralement très enthousiaste à l'égard de l'IA (Dandurand et al., 2023), a aussitôt suscité des débats quant à l'usage des IA dites génératives et ses implications pour le journalisme et l'information sous toutes ses formes (texte, audio, images fixes, vidéo).

La mésinformation et la désinformation ont été classés comme le premier risque mondial pour les deux prochaines années par le World Economic Forum (2024), notamment en raison des nombreuses élections qui auront lieu durant cette période. Selon ce même rapport, l'enjeu demeurera parmi les principaux risques mondiaux pour les prochaines années, de même que les effets néfastes de l'intelligence artificielle. Il s'agit d'un problème transversal qui concerne tout particulièrement les médias d'information (Fallaha et al., 2024 : p. 21).

De nombreux médias, associations et autres organisations des milieux journalistiques se sont dotés ces derniers mois de chartes d'IA, la plupart assez générales et préliminaires. L'analyse de 20 chartes principalement européennes[14] a permis de dégager certains principes communs (Pettineo, 2024),dont :

- L'utilisation de l'IA comme un appui aux capacités humaines et non un substitut;
- Le contrôle éditorial strict sur l'utilisation des outils et systèmes d'IA;
- La vérification systématique par des journalistes des contenus avant utilisation ou diffusion;
- La formation continue des journalistes et le recrutement de professionnels compétents;
- La transparence à l'égard des utilisateurs et le respect de l'intégrité de l'information;
- Le respect du droit d'auteur;
- La participation des médias à la gouvernance et à la réglementation de l'IA;
- L'application générale de la déontologie journalistique;
- La transparence et traçabilité de la production IA.

Au cœur de cette réflexion, celle de la confiance du public à l'égard des médias, déjà durement éprouvée, en particulier depuis la pandémie de COVID-19 (Brin et al., 2023). En outre, la capacité d'adaptation, de développement et de rétention d'expertise en IA au sein des médias, surtout dans des petits marchés comme le Québec, semble limitée (Académie de la transformation numérique et Centre d'études sur les médias, 2023[15]).

12 Voir notamment sur ce point : Clémence Varin, « La prise en compte de l'impact environnemental de la protection et la promotion de la diversité des expressions culturelles dans l'environnement numérique par les Parties à la Convention de 2005 : un état des lieux » dans Penser le lien culture-nature en droit : réflexions, réalisations, aspirations, PUL, 2023 Voir aussi : Holzapfel, A. (2022). « Introducing Political Ecology of Creative-Ai »

13 Voir le reportage de Radio-Canada sur le projet Channel One, diffusé le 29 avril 2024. https://ici.radio-canada.ca/info/videos/1-8930498/bulletins-televises-generes-par-intelligence-artificielle

14 Incluant également la CBC.

15 Cette collaboration entre l'ATN et le CEM inclut quatre documents (une synthèse des résultats et trois rapports de recherche).

## L'IA et l'accès à la culture et à l'information

### Culture

La création n'est pas le seul maillon de la chaîne de valeur culturelle touché par l'IA. L'accès aux expressions culturelles dans l'environnement numérique est aussi bouleversé par l'essor de cette technologie dans le domaine des industries culturelles et créatives (Brin et Guèvremont, 2024). Le recours des grandes plateformes numériques à l'IA et aux algorithmes de recommandation a une forte répercussion sur l'accès à certains contenus culturels nationaux et locaux, notamment lorsque ces contenus ne sont pas des nouveautés[16]. Le sujet n'est pas entièrement nouveau, mais les préoccupations relatives à l'impact de cette utilisation de l'IA dans la prescription algorithmique ne font que s'accroître. L'objectif premier de ces grandes entreprises multinationales est de mettre en avant leurs propres contenus (dits « contenus originaux »), ce qui tend à appauvrir la diversité des expressions culturelles réellement accessibles et visibles sur les plateformes de diffusion de contenus. Ces enjeux de découvrabilité[17] des contenus culturels concernent aussi l'accès à des contenus dans certaines langues, par exemple des contenus d'expression originale de langue française, souvent défavorisés par les algorithmes de recommandation (Beaudoin et al., 2024 : p. 15-16)[18]. Cela pose donc la question du droit à l'accès à sa propre culture, indissociable de la préservation des identités.

Enfin, l'utilisation des systèmes d'IA dans le domaine du patrimoine culturel engendre aussi certaines inquiétudes. Certes, l'IA peut être source d'opportunités, par exemple lorsqu'elle est utilisée pour numériser, préserver et rendre accessibles des œuvres fragiles telles que des peintures ou manuscrits anciens, (Culture Montérégie, 2024). Cependant, la numérisation de certains types de patrimoine soulève aussi des risques d'appropriation culturelle et de non-respect de la confidentialité de certaines données (Culture Montérégie, 2024). Des enjeux éthiques émergent aussi de la numérisation du patrimoine culturel immatériel, un patrimoine considéré indissociable des communautés qui le portent et en assurent généralement la transmission; la mise à disposition de ce patrimoine dans l'environnement numérique ne risque-t-elle pas de figer un patrimoine que l'on qualifie généralement de « vivant »?

Finalement, l'utilisation de l'IA pour traduire automatiquement des textes séduit autant qu'elle inquiète : tout en permettant de diffuser plus largement un patrimoine documentaire et littéraire, cette utilisation soulève de nombreuses questions sur la capacité des systèmes d'IA de restituer fidèlement un œuvre dans son contexte, ou encore dans le respect des subtilités souvent inhérentes à une œuvre littéraire.

### Médias

Au Canada, le blocage du contenu des médias d'information sur les plateformes du groupe Meta à la suite de l'adoption de la loi C-18 a rappelé l'influence considérable des géants du numérique sur la visibilité des contenus journalistiques (Parker et al., 2024). En effet, les médias sociaux sont le principal moyen d'accès aux nouvelles pour 1 Canadien sur 4 et près de la moitié des 18-34 ans (Tremblay et Charlton, 2023). Par ailleurs, les moteurs de recherche comme Bing ou Google intègrent des outils d'IA directement à leurs moteurs de recherche permettant aux utilisateurs d'obtenir rapidement une information avec une simple requête. Ces acteurs, grâce à leur position dominante et leur expertise de pointe en matière d'IA, pourraient influencer la direction du trafic sur les navigateurs au détriment des médias d'information (Simon et Isaza-Ibarra, 2023).

L'accès à une information de qualité constitue une seconde préoccupation. La crise du journalisme pousse les entreprises de presse à revoir leurs modèles économiques et l'introduction des outils d'IA pourrait accentuer cette tendance. Dans un contexte où l'abonnement payant tend à s'imposer comme modèle d'affaires pour les médias écrits (Cardon, 2019)[19], les technologies d'IA représentent un enjeu pour l'accès à une information de qualité dans la mesure où elles permettent la diffusion et la création massive de contenus manipulés (texte, photo, voix, vidéo). Il est désormais à la portée d'un utilisateur non expert de créer un site web de désinformation avec des outils d'IA générative. Pour le citoyen, distinguer le vrai du faux devient alors d'autant plus complexe (Labrecque, 2023).

Enfin, sur une note plus positive, les technologies d'IA permettent aux journalistes d'accéder, de synthétiser et de fournir au public un plus grande quantité d'informations, grâce notamment à la capacité d'extraction et d'analyse rapide de données massives, y compris en journalisme d'investigation (Stray, 2021).

16 À ce sujet, voir notamment cet article publié dans Le Devoir : https://www.ledevoir.com/culture/cinema/811460/cinema-plusieurs-films-quebecois-quasi-introuvables

17 La découvrabilité se définit par le potentiel pour un contenu, disponible en ligne, d'être aisément découvert par des internautes dans le cyberespace, notamment par ceux qui ne cherchaient pas précisément le contenu en question. Source : https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26541675/decouvrabilite

18 À ce propos, les langues des signes sont souvent oubliées dans les réflexions sur la découvrabilité des contenus culturels en langues minoritaires. Or, bien qu'ils ne s'agissent pas de langues dites parlées, il demeure important de ne pas les omettre, particulièrement lorsque ces réflexions évoquent un accès large et inclusif à une diversité d'expressions culturelles.

19 Au Québec, *La Presse* et les quotidiens de Québecor (*Le Journal de Montréal/Le Journal de Québec*) privilégient toujours la gratuité en ligne. Les six quotidiens régionaux des Coops de l'information (CN2I) ont érigé un mur payant en 2020 (Giroux, 2022 : 127).

# Débats scientifiques

## Culture

L'utilisation de l'IA dans le domaine des industries culturelles et créatives est source de nombreux débats. Le plus important est sans doute l'impact de l'IA sur la créativité, que certains envisagent d'abord sous l'angle des risques, alors que d'autres préfèrent explorer le sujet sous l'angle des opportunités. Selon certains, le droit d'auteur a notamment pour finalité d'encourager la créativité par l'attribution de droits exclusifs aux créateurs; ainsi, en s'abreuvant d'œuvres sans autorisation des titulaires de droits, l'IA constituerait ainsi une menace pour la juste rémunération des artistes et, par le fait même, pour la créativité humaine. Pour d'autres, l'IA serait plutôt source de créativité en augmentant le potentiel du créateur (Kulesz parle d'artiste « augmenté »[20]), ou encore la productivité des industries culturelles et créatives. La production de films, de musiques ou encore d'images à moindre coût pourrait en outre enrichir l'offre de contenus culturels et rendre ces dernières plus facilement accessibles, parce que moins coûteux. Mais une sur-offre n'est pas sans danger : certains auteurs craignent notamment le phénomène d'auto-génération de l'IA : plus des contenus seront créés par IA, plus ces systèmes seront amenés à s'alimenter de contenus similaires, créant ainsi un environnement stérile qui amènerait l'IA à s'étouffer elle-même (Arielli et Manovich, 2024). L'IA serait-elle alors vraiment source de créativité?

Un autre débat porte sur la juste rémunération des artistes dont les œuvres servent à entraîner les systèmes d'IA. En effet, la plupart des entreprises entraînent leurs modèles d'IA sans le consentement des créateurs qui non seulement ne semblent pas avoir voix au chapitre, mais surtout, ne bénéficient d'aucune compensation pour un tel usage. Pour certains, le droit d'auteur en vigueur n'est pas adapté aux IA génératives qui, par nature, violent le droit d'auteur en se nourrissant d'œuvres sans autorisation préalable; par conséquent, une réforme s'impose. D'autres souhaitent éviter une révision du droit en vigueur (Rioux, 2024), craignant peut-être qu'une révision des règles engendre davantage un affaiblissement qu'un renforcement de la protection offerte aux auteurs. Certains auteurs proposent par ailleurs des modèles alternatifs de rémunération, qui s'appuierait par exemple sur un système de prélèvement orienté, imposant une obligation générale de contribution financière de la part des fournisseurs de systèmes d'IA générative. Les sommes ainsi prélevées pourraient ensuite être distribuées aux auteurs et autres titulaires de droit; elles pourraient alors être versées dans des fonds sociaux et culturels (Senftleben, 2023).

## Médias

Les IA génératives, en particulier celles basées sur les larges modèles de langage, sont considérées comme une perturbation potentielle majeure dans le domaine du journalisme (Arguedas et Simon, 2023). L'utilisation d'outils d'IA permet de personnaliser l'offre de contenus, mais aussi d'orienter la production des contenus vers des catégories particulières d'usagers. Ceci pose d'une nouvelle manière le vieux problème du « public imaginé » en journalisme, vision réductrice ou biaisée du public réel dans sa multiplicité et sa complexité (Siegel 2024). Pour certains, l'introduction de processus automatisés confinerait le public à une consommation passive (Béasse 2021) alors que d'autres y voient le potentiel de rapprocher l'information des citoyens et de favoriser la diversité des perspectives (Lin et Lewis 2022). En outre, il n'est pas acquis que le public favorisera le travail journalistique « humain » ou plutôt les systèmes d'IA, jugés plus neutres et rigoureux (Jung, 2023).

La question du respect des droits d'auteur se pose aussi avec acuité pour les médias. Les modèles d'IA générative ont été développés grâce au moissonnage d'un grand nombre de données de sources nombreuses, y compris celles issues de contenus journalistiques. Le cas de la poursuite en justice engagée par *The New York Times* à l'encontre d'OpenAI et Microsoft illustre une opposition à l'utilisation gratuite de contenus journalistiques pour entraîner les modèles de robots conversationnels[21]. À l'inverse, le média d'information français *Le Monde* a conclu un accord de partenariat avec OpenAI, autorisant l'entreprise à utiliser le contenu du média pour entraîner ChatGPT en échange d'une compensation financière[22]. Néanmoins, ce genre d'accord soulève des questions sur les rapports de pouvoir entre les médias d'information et les géants des technologies de l'IA (Faljaoui, 2024).

20 Voir notamment ce rapport de l'OIF de 2019 « L'intelligence artificielle dans l'art et les industries culturelles et créatives. Panorama des technologies, expertises et bonnes pratiques dans l'espace francophone » à la page 33 : https://www.francophonie.org/sites/default/files/2020-05/OIF_Guide-IA-art_VF.pdf

21 Cet article du *New York Times* offre des détails supplémentaires : https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

22 *Le Monde* justifie cet accord par la volonté de protéger le modèle économique du média et de protéger son identité éditoriale en mettant en place une charte d'éthique et de déontologie spécifique à l'utilisation de l'IA. Pour en savoir plus : https://www.lemonde.fr/le-monde-et-vous/article/2024/03/13/intelligence-artificielle-un-accord-de-partenariat-entre-le-monde-et-openai_6221836_6065879.html À noter que Prisa (Espagne), Associated Press (États-Unis), Axel Springer (Allemagne) et le *Financial Times* (Royaume-Uni), ont aussi conclu des accords avec OpenAI.

# Cas d'usage

## Fairly Trained

Fairly Trained est un organisme de certification ayant pour objectif d'offrir une certification aux fournisseurs de services d'IA qui n'utilisent aucune œuvre protégée par des droits d'auteur sans licence. Cet organisme a été lancé par le fondateur d'un outil d'intelligence artificielle qui s'est rendu compte que les bases de données utilisées pour entraîner l'outil avaient repris le travail d'artistes sans leur autorisation. Cette certification peut être octroyée aussi bien à des entreprises, que des organisations ou même des produits. Cet exemple illustre ce qu'on appelle le « FAIR Data ». Contrairement à l'Open Data, le FAIR Data n'exige pas une gratuité ou un libre accès total mais exige toutefois la clarté et la transparence des conditions d'accès et de réutilisation. Contrairement au concept général et ambigu « open », l'accessibilité transparente mais contrôlée - FAIR - des données et des services permet la participation d'un large éventail de secteurs (publics et privés) ainsi qu'un véritable partenariat égal avec les parties prenantes dans toutes les sociétés du monde (Barend et al. 2017).

## Sora

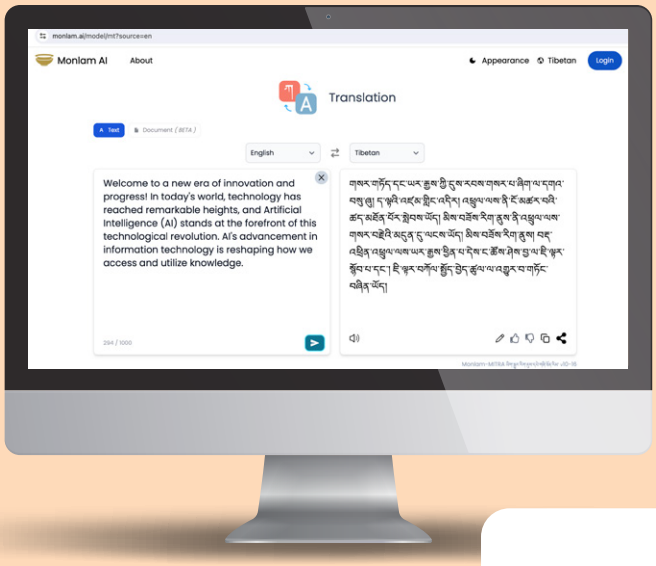
Sora est un logiciel générateur de vidéos créé par OpenAI qui permet de générer des vidéos à partir d'une simple ligne de texte. Visuellement très impressionnante, la qualité des vidéos générées par Sora serait « digne d'Hollywood » selon certains (Metz, 2024), qualité qui effraie les artistes du milieu du cinéma. OpenAI marque déjà les vidéos produites par Sora de filigranes les identifiant comme générées par l'intelligence artificielle mais l'entreprise a également reconnu que ces filigranes peuvent être supprimés facilement et peuvent être difficiles à repérer (Metz, 2024). Sora représente certes un potentiel créatif illimité mais elle inspire également beaucoup de crainte aux artistes dans la mesure où déceler les créations réalisées par l'IA semble devenir de plus en plus ardu (Agence France-Presse, 2024).

## Juice

Lancée en 2022 par l'entrepreneur français Rafi Haladjian, l'application mobile Juice offre un service audio d'« info augmentée » sous une forme personnalisée, animée par des voix de synthèse et alimentée par de multiples producteurs de contenus (radio, baladodiffusion, musique...). Juice s'inscrit dans la continuité des initiatives de ce pionnier du numérique et des objets connectés. L'application fonctionne comme une sorte d'assistant qui connaît et anticipe les intérêts et préférences de l'usager, y compris ses opinions politiques. Elle fonctionne entièrement sur commande vocale et sans écran (Radio France, 2022).

## Monlam AI

Monlam AI est un outil d'intelligence artificielle visant à préserver le vaste patrimoine culturel tibétain, notamment la littérature, l'histoire, la musique et les textes bouddhistes. Cet outil intervient alors que la Chine restreint de plus en plus les droits linguistiques et culturels au Tibet. En effet, face aux efforts croissants de la Chine pour éroder la langue tibétaine, notamment en interdisant l'enseignement de la langue tibétaine dans les écoles de certaines régions, un tel outil se révèle être « a significant step forward for the Tibetan community, as it embraces modern technology to preserve its cultural heritage and facilitate communication in the digital age » (RFA, 2023).



# Pistes d'action et implications pour les parties prenantes

## 1

### Culture

- Documenter l'impact réel de l'IA sur les artistes et autres professionnels de la culture, ainsi que sur les industries culturelles et créatives du Québec.
- Identifier les besoins de ces acteurs en termes de renforcement de leur capacité afin de leur permettre de tirer tous les bénéfices potentiels du recours à l'IA dans leurs pratiques et activités respectives.
- Initier une réflexion sur le renforcement du statut de l'artiste à l'ère de l'IA, en précisant notamment les notions de juste rémunération ou de juste compensation.
- Développer une réflexion approfondie sur l'encadrement juridique du recours aux systèmes d'IA dans le champ des industries culturelles et créatives, en identifiant les principes fondamentaux qui pourraient être inscrits dans de nouvelles normes en la matière, par exemple le principe de transparence.
- Poursuivre les recherches sur les mesures, législatives ou d'autre nature, qui devraient être déployées en vue de stimuler la découvrabilité des contenus culturels nationaux, locaux et dans certaines langues sur les plateformes, mais aussi par le biais des outils conversationnels en plein essor.

## 2

### Médias

- Formation initiale et continue des journalistes à l'IA (usage critique; collaboration avec spécialistes).
- Analyse des niveaux de risque de différentes formes d'IA selon le contexte (p.ex. protection de la confidentialité des sources en journalisme d'investigation).
- Développement de bonnes pratiques journalistiques et de chartes IA par les médias, associations et instances d'autorégulation.
- Développement d'une culture de l'innovation, attraction et rétention d'expertises au sein des médias.
- Éducation du public à l'IA : évaluation et détection des contenus générés par IA.

# Axe
# Sobriété numérique et transition socio-écologique

**Sous la responsabilité de :**
Christophe Abrassart et Stéphane Roche

**En collaboration avec :**
Florent Begue et Alexandre Gourret

Le numérique et l'environnement ont une relation ambivalente. D'un côté le numérique peut contribuer, à certaines conditions, au développement d'outils de mesure et de gestion de la transition socio-écologique. D'un autre côté, de plus en plus de rapports nous alertent sur l'empreinte environnementale croissante du numérique et de l'intelligence artificielle. C'est sur ce second volet que l'axe Sobriété numérique et transition socio-écologique a choisi d'enquêter cette année. Cette note présente dans un premier temps la tendance actuelle et les projections prospectives sur l'empreinte environnementale du numérique. Elle montre ensuite que deux grandes stratégies sont actuellement défendues par les experts face à cet enjeu : celle de l'écoconception des équipements et services numériques et celle de la sobriété numérique. Les termes de ce débat et ses liens avec les enjeux sociaux du numérique sont précisés et représentés à l'aide d'une cartographie visuelle. Cette note présente enfin un cas d'usage de sobriété numérique et la démarche participative qui sera mise en œuvre avec les parties prenantes à l'aide de scénarios de sobriété numérique au Québec en 2050.

# Actualité scientifique

## L'augmentation de l'empreinte environnementale du numérique au Québec et dans le monde

Le numérique et ses développements contemporains, comme l'intelligence artificielle, l'internet des objets, les téléphones intelligents , le streaming, la réalité virtuelle et augmentée, se diffusent de plus en plus dans la société. En conséquence, leur empreinte environnementale ne cesse d'augmenter. D'après une récente étude de référence produite par l'Agence de la transition écologique (ADEME) et l'Autorité de régulation des communications électroniques, des postes et de la distribution de la presse (Arcep), à l'échelle d'un pays comme la France, cette empreinte représentait en 2020 10 % de la consommation électrique du pays et 2,5 % des émissions de GES (ADEME-Arcep, 2023). À l'échelle mondiale, les estimations sont autour de 10 % de l'électricité et entre 3 et 4 % des GES en 2018 (Shift Project, 2021).

Un des résultats de cette étude est que l'empreinte carbone du numérique dépend essentiellement des équipements et de leur fabrication : en 2020, **par composantes du numérique**, 79 % de l'impact provient de la fabrication et de l'utilisation des équipements grand public (téléphones intelligents, téléviseurs…), 5 % des réseaux fixes et mobiles, et 16 % des centres de données.

Toujours en 2020, **par phase du cycle de vie**, 78 % de l'empreinte carbone provient de l'extraction des matières premières et de la fabrication des équipements grand public, des centres de données et des réseaux, 1 % de l'empreinte carbone est lié à la distribution, et 21 % est liée à l'utilisation des équipements, des réseaux et des centres de données.

## L'empreinte environnementale du numérique en chiffres

1 courriel correspond à 4 g $CO_2$ éq.; et à 35 g avec une pièce jointe $CO_2$ Éq./unité (ADEME, s.d.). La visite d'un site Internet peut émettre entre 1 et 5 g $CO_2$ éq., selon le poids de la page consultée (EcoIndex, s.d.). Du côté des équipements, la taille des écrans est une variable clé : la fabrication d'un ordinateur portable émet 156 Kg $CO_2$ éq. et demande 588 Kg de matières premières; la fabrication d'un téléphone intelligent de petite taille émet 16,5 Kg $CO_2$ éq. (ADEME, s.d.). Dans le monde, 62 millions de tonnes de déchets électroniques ont été produits en 2022 (projection de +32 % d'ici 2030), soit 7,8 Kg en moyenne par habitant, avec de fortes disparités : 14,1 Kg/hab dans les Amériques, contre 6,4 Kg/hab en Asie selon le Global E-Waste Monitor (UNITAR, 2024).

D'autres enjeux environnementaux et sociaux peuvent être mentionnés, comme la destruction de la biodiversité et le non respects des droits humains dans les activités minières pour extraire les métaux nécessaires à la fabrication des équipements numériques, le détournement local de la ressource en eau pour refroidir les centres de données, ou les enjeux de pollution liés à l'exportation des déchets électroniques dans certains pays du Sud.

## Une tendance qui devrait s'accélérer avec les IA génératives

Cette tendance pourrait de plus s'amplifier avec le déploiement à grande échelle des IA génératives, pour lesquelles des premières estimations commencent à être établies malgré le peu de données publiées par les entreprises. ChatGPT a demandé plusieurs jours d'entraînement sur des milliers de GPU ce qui a consommé 1287 MWh d'électricité et conduit à des émissions d'environ 552 tonnes $CO_2$ éq. (Patterson et al., 2022). De plus, l'utilisation quotidienne d'IA génératives par des millions de personnes serait beaucoup plus gourmande en électricité. Ainsi Luccioni et al. (2023) estime que l'empreinte carbone de la génération d'images ou de textes peut être significative, tout en variant d'un modèle à un autre : générer 1.000 images peut consommer jusqu'à 2,9 KWh d'électricité et générer 1000 textes jusqu'à 0,047 KWh (une recharge de téléphone intelligent = environ 0.015 kWh).

## Des projections en 2050 incompatibles avec les limites planétaires.

L'empreinte du numérique pourrait croître d'ici 2050 avec une augmentation de près de +200 % pour les GES et près de +100 % pour la consommation électrique en France, avec des tensions sur la disponibilité de certains métaux entrant dans la composition des équipements (ADEME-Arcep, 2023). Les facteurs déterminants de ces projections sont notamment le degré de diffusion de l'écoconception et de l'économie circulaire sur les équipements, centres de données et réseaux, la maîtrise des effets rebonds et de l'obsolescence programmée sur les terminaux, mais aussi les mesures de sobriété individuelles et collectives dans les usages. En effet, dans ces scénarios à l'horizon 2050, seuls ceux qui prévoient des mesures de sobriété fortes dans la société permettent d'éviter ces augmentations.

L'empreinte carbone du numérique se situe autour de 3 à 4 % des émissions GES, soit environ 2Gt $CO_2$ éq. (Shift Project 2021). En ordre de grandeur, avec une croissance de 6 à 8 % par an, cette empreinte du numérique pourrait dépasser 10 Gt $CO_2$ éq. en 2050, ce qui correspond au budget mondial de GES à atteindre à cette date pour rester sur une trajectoire de 1.5 ou 1.8 degrés (entre 10 et 20 Gt éq. $CO_2$ (IPCC 2022)). Ramené à une personne, ce budget revient à 1 à 2 t $CO_2$ éq. par habitant par an. Un budget de carbone si restreint nous place devant un sérieux dilemme : quel numérique garderons-nous en 2050 face aux usages essentiels (alimentation, mobilité, habitat, santé)? Devrons-nous renoncer en partie au numérique ou à l'objectif de deux degrés en 2100?

# Débat scientifique

## Écoconception versus sobriété du numérique

Deux grandes stratégies sont défendues par les experts et parties prenantes et reprises dans des politiques publiques environnementales : celle de l'écoconception des équipements et des services numériques et celle de la sobriété numérique. La perspective adoptée ici est celle d'une enquête sur les instruments de gouvernement et leurs effets de cadrage et de problématisation (Lascoumes et Le Galès, 2005). L'écoconception et la sobriété conduisent en effet à deux cadrages différents du problème : réduire les impacts environnementaux tout au long du cycle de vie d'un produit pour la première, et repenser de manière critique et créative nos usages et nos styles de vie pour la seconde.

## Les stratégies d'écoconception, leurs apports et leurs limites

Selon la norme 14006 de l'organisation internationale de normalisation (ISO), l'éco-conception est une « approche méthodique qui prend en considération les aspects environnementaux du processus de conception et développement dans le but de réduire les impacts environnementaux négatifs tout au long du cycle de vie d'un produit » (ISO, 2020). Cette démarche, qui est la plus diffusée, a pris son essor dans les années 2000 chez les ingénieurs et les designers. Différents régimes d'écoconception sont possibles (Abrassart, 2011). Tout d'abord **l'éco-conception réglée**, qui vise à optimiser les équipements, centres de données et services numériques existants à partir de méthodes comme l'analyse du cycle de vie et des règles d'écoconception, mais sans changer l'identité de ces objets (leurs fonctions, leurs scénarios d'usage). Plusieurs référentiels d'écoconception sont parus ces dernières années, notamment le ***Référentiel général des services numériques*** de l'Arcep-Arcom (Arcep, Arcom et ADEME, 2024), le plus récent, qui inclut également des critères sur les systèmes d'IA. Le second régime d'écoconception est celui de **l'écoconception technico-innovante**, qui passe par une réflexion créative sur les paradigmes de performance et de valeur des objets, mais sans bouleverser leurs usages. Par exemple : un téléphone circulaire comme le FairPhone, l'IA frugale avec beaucoup moins de données, un robot permettant de recycler un téléphone intelligent, un centre de données immergé dans l'océan pour son refroidissement.

Une limite de l'écoconception est l'effet rebond, par lequel une plus grande efficacité énergétique et une meilleure performance environnementale des équipements et services numériques conduit à une augmentation plus que proportionnelle de leur usage. Par exemple, on constate une efficacité énergétique accrue des centres de données, mais aussi l'explosion du stockage des données dans le cloud. La recyclabilité des téléphones intelligents augmente mais leurs écrans sont de plus grande taille.

Plusieurs référentiels d'écoconception proposent un questionnement préalable et critique sur la nécessité du besoin de numérisation (exemple : Designers éthiques (2022)). Cette interrogation sur la pertinence du projet relève du principe de sobriété.

## Une autre approche : la sobriété, un concept au cœur de politiques publiques contemporaines

Certains experts pensent en effet que ces démarches techniques d'écoconception, même si elles sont incontournables, ne suffiront pas à réduire les impacts environnementaux du numérique et proposent une autre approche : la sobriété numérique (Shift Project, 2020; ADEME-Arcep, 2023). D'après un rapport de l'ADEME (Cézard et Mourad, 2019), la sobriété est décrite comme une approche systématique intégrant la modération dans la consommation et la production de ressources. Elle invite à une remise en cause des usages et des comportements individuels et collectifs (Cézard et Mourad, 2019; Flipo, 2017; Villalba, 2022). Pour le rapport de l'IPCC (2022, p. 31) la sobriété renvoie aux « politiques de *sufficiency* » qui rassemblent « *a set of measures and daily practices that avoid demand for energy, materials, land and water while delivering human well-being for all within planetary boundaries* ».

Depuis quelques années on assiste ainsi à la mise à l'agenda du concept de sobriété dans différents champs de politiques publiques : des exercices de prospective gouvernementaux sur la transition incluent maintenant un scénario de sobriété (ex. le scénario « Sobriété » dans RTE 2050 sur la transition énergétique en France à l'horizon 2050; ou le scénario « Génération frugale » de l'ADEME (2024) sur la transition bas carbone en France à l'horizon 2050). De plus, des politiques publiques thématiques de sobriété ont vu le jour récemment, par exemple sur la sobriété foncière, pour lutter contre l'étalement urbain et l'artificialisation des sols (dispositif « zéro artificialisation nette »); sur la sobriété matérielle : l'interdiction de l'obsolescence programmée en France et au Québec (Loi 29). Un défi central de ces politiques est de rendre la sobriété socialement désirable, par des démarches de co-construction collective à l'aide de mises en récit stimulantes.

S'intéresser à la sobriété c'est aussi aborder de manière indissociable des enjeux éthiques et sociaux du numérique comme la capture de l'attention par des « nudges », l'exposition des enfants et adolescents aux écrans, la protection des données personnelles et des libertés publiques (LINC, 2023), ou encore les risques des fausses nouvelles pour la démocratie. C'est également expliciter les questions de priorisation des usages du numérique dans un monde aux ressources finies (ex. le numérique pour l'écologie et la santé ou pour les loisirs et le marketing expérientiel?).

Un budget de carbone si restreint nous place devant un sérieux dilemme : quel numérique garderons-nous en 2050 face aux usages essentiels (alimentation, mobilité, habitat, santé)? Devrons-nous renoncer en partie au numérique ou à l'objectif de deux degrés en 2100?

## Une cartographie des cadrages et problématisations

Le schéma ci-dessous propose une cartographie de ces deux stratégies en quatre quadrants. Deux polarités sont considérées : la première oppose une logique d'optimisation du système sociotechnique existant à une autre de transformation de ce système. La seconde polarité oppose une recherche d'efficacité de fonctionnement sans questionner les fonctions d'usage (écoconception) à une démarche de reconception critique et créative des fonctions d'usage (sobriété). Au croisement de ces deux polarités apparaissent quatre quadrants permettant de thématiser quatre idéotypes de stratégies de durabilité dans le numérique. Pour illustrer cette cartographie, prenons **l'exemple des écrans numériques de grande taille**, dont l'empreinte environnementale est significative, et dont le rythme de renouvellement est guidé par l'obsolescence programmée technique et logicielle.

Schéma : Abrassart C. et al., ACFAS 2024

L'approche dominante est de faire de **l'écoconception réglée** des écrans en suivant des cahiers de charges techniques pour optimiser leur fonctionnement (en bas à gauche). Pour les écrans les plus efficients, et dont la taille est appropriée à l'usage et la durée de vie allongée, un écolabel peut être obtenu (ex. Energy Star au Canada ou aux États-Unis, l'Écolabel européen en Europe). Ensuite une approche **d'écoconception technico-innovante** peut être mobilisée, par exemple en adoptant le principe d'économie de la fonctionnalité, qui consiste à vendre un service plutôt qu'un produit (en bas à droite). On pourrait par exemple imaginer une solution de partage qui s'appellerait « Commun'Écran », sur le modèle de Communauto, avec une réparation continue de ces équipements dans une logique d'économie circulaire. Ou encore un vidéoprojecteur de poche hyper efficient partagé entre voisins. Une troisième approche est celle de la **simplicité volontaire** soit celle de la sobriété mais sans changer le système actuel (en haut à gauche). Elle peut conduire à renoncer à s'équiper d'un écran à la maison et à ne pas regarder de séries en streaming depuis chez soi, et préférer aller au cinéma ou au théâtre, ou même intégrer une troupe de théâtre pour apprendre à jouer, dans une logique de « suffisance intensive » (Monnin et Ben Kemoun, 2022). Enfin, nous proposons un quatrième quadrant, celui de la **sobriété socialement innovante du numérique**, qui vise la reconception des fonctions d'usage et contribue à la transformation du système. Par exemple se regrouper dans une bibliothèque tiers-lieu pour participer au développement de communs de vidéos en très basse résolution, selon un nouveau paradigme esthétique, ou pour des soirées de projection basse résolution (ex. « low-tech numérique » (Roussilhe et Nova, 2020), ou approche « Permanumérique »).

# Cas d'usage

## Le VéLI, la mobilité en sobriété numérique innovante

Le secteur des transports est la principale source d'émissions de GES au Québec, avec 43 % des émissions en 2021. Et les solutions envisagées de mobilité durable comme l'électrification, l'autopartage ou la multimodalité autour des transports en commun ne conduisent pas à une baisse significative de ces émissions, notamment à cause de l'augmentation tendancielle de la taille des voitures. Une approche de sobriété numérique socialement innovante, permettrait-elle d'imaginer une autre trajectoire de transition dans les transports au Québec?

Un scénario prospectif repose sur une nouvelle identité des « véhicules », non pas vers plus d'autonomie, mais vers plus de sobriété et une taille plus petite. Ces **véhicules légers intermédiaires (« VÉLI »)** sont définis comme tout mode de transport individuel de moins de 600 kg, compris entre le vélo et la voiture classique » (Bigo et al., 2021). Ces petits véhicules, propulsés par le pédalage combiné à une petite puissance électrique comparable à celle des vélo-cargos électriques, sont aussi regroupés sous le terme émergent de **minimobilité** (Riggs et Shukla, 2022). Dans ce scénario, un « numérique juste suffisant » pourrait jouer un rôle déterminant pour la mise en circulation sécuritaire et désirable des VéLIs au Québec : pour la navigation sur le territoire et la connection avec les gares de train ou de bus, pour organiser des plateformes de « VÉLI-partage », pour la maintenance des VéLIs en économie circulaire, ou encore pour assurer la sécurité des VÉLIs sur la route à l'aide de capteurs et de systèmes intelligents de régulation de la vitesse.

# Pistes d'action et implications pour les parties prenantes

**Des scénarios sur la sobriété numérique socialement innovants au Québec en 2050, seront prochainement produits et mis en débat avec un large public dans le cadre d'ateliers d'experts et de conversations citoyennes, afin de générer des apprentissages pour mieux définir des politiques potentielles de sobriété numérique au Québec. Ces scénarios mettront en valeur :**

1

## Des dispositifs de sobriété variés, par exemple une priorisation de l'usage des données et des algorithmes dans certains secteurs essentiels

2

## Des logiques de regroupement spatial pour l'usage du numérique

3

## Des équipements dont l'usage est grandement prolongé par la maintenance et la réparation dans une logique de care et de « circularité forte »

(Aggeri et al., 2023)

4

## Des nouvelles politiques de mémoire et d'effacement sélective des données avec des centres de données ayant le statut de « biens communs mondiaux »

# Axe

# Droit, cyberjustice et cybersécurité

**Sous la responsabilité de :**
Vincent Gautrais et Anne-Sophie Hulin

**En collaboration avec :**
William Auclair

L'intelligence artificielle suscite passion et émotion. Les questions de gouvernance en lien avec son encadrement sont susceptibles d'offrir, tant par sa substance que par les institutions intéressées, un cadre capable de penser « froidement » la matière. Ce court texte tente, au-delà du constat des changements tant dans les faits que dans la manière de les encadrer, d'identifier des voies possibles pour mieux appréhender ce futur (déjà) très proche.

# Actualité scientifique

## L'IA : cas d'urgence normative?

Depuis l'onde de choc ChatGPT, l'encadrement de l'intelligence artificielle (IA) est pris d'un vif sentiment d'urgence qui s'explique par la conjonction de deux facteurs : d'une part, le rythme effréné auquel se réalisent les progrès technologiques en ce domaine et, d'autre part, la prise de conscience collective de la nature et de l'étendue des bouleversements sociétaux engendrés par l'expansion de l'IA (notamment générative) au sein de la société. Ce sentiment d'urgence a suscité une diversité de réactions. Certains dans la communauté de recherche et dans l'industrie ont, par exemple, fait valoir la nécessité d'un temps d'arrêt de la recherche et développement (RetD) en IA. Tandis que plusieurs régulateurs se sont engagés, avec frénésie et ambition, dans une course normative pour davantage encadrer l'IA tant au niveau national qu'international, d'autres alertent l'opinion publique sur les effets d'une réglementation qui, adoptée par réaction et dans la précipitation, comporte le risque de choix normatifs hâtifs et non pérennes. Ce sentiment d'urgence a également eu pour effet indirect de passionner le débat relatif à l'encadrement de l'IA et d'en accentuer la polarisation – certes déjà existante – entre la promotion de l'innovation technologique, essentielle au développement économique et à l'avancement des connaissances, et la nécessité de protéger les droits humains et sociétaux. Les positions étant extrêmement tranchées, l'établissement de compromis semble de plus en plus difficile à atteindre, au point de faire émerger un risque d'adoption de mesures disproportionnées ou non équilibrées.

## Option interventionniste

Ce faisant, cette polarisation donne naissance à un schisme normatif entre, d'un côté, ceux qui plaident pour une réglementation stricte et contraignante, arguant que sans garde-fous légaux rigoureux, les risques associés à l'IA pourraient surpasser ses bénéfices, et, d'autre part, ceux qui prônent une approche plus souple, favorisant l'autorégulation et les principes éthiques volontaires pour ne pas entraver l'innovation. Fortes de ce mélange « urgence, passion et polarisation », les deux dernières années ont été le théâtre d'un tournant marquant du point de vue de la gouvernance de l'IA : sous l'effet d'une offensive juridique qui emboîte le pas à l'éthique, un encadrement de l'IA avec davantage de mordant est mis de l'avant. Une gouvernance qui, tout en intégrant cette quête de davantage de prise de contrôle, présente un spectre des possibles d'une grande variété. Comment opérer ce resserrement des pratiques? Quels sont les acteurs, notamment étatiques, qui doivent intervenir dans cette reprise en main?

## Comment?

Cette juridicité montante va de pair avec la consolidation de modèles de gouvernance de l'IA, que l'on peut subdiviser de la sorte : **d'une part**, des modèles qui prennent corps autour d'un ou de plusieurs textes juridiques spécifiques à l'IA et, **d'autre part**, des modèles pour lesquels le droit « dur » ne saurait être l'épicentre de la gouvernance de l'IA. Cette dichotomie n'est évidemment pas absolue et connaît des formes intermédiaires dont la diversité illustre la complexité de la tâche. Dans cette perspective, le modèle européen constitue actuellement le paroxysme de la juridicisation de la gouvernance de l'IA avec le *Règlement sur l'intelligence artificielle* (Parlement européen, 2024). Reposant sur une approche fondée sur les risques, cette loi de l'Union européenne dresse une liste de pratiques interdites en raison d'un niveau de risque jugé inacceptable et pose des obligations strictes pour la mise en marché des systèmes présentant un niveau de risque élevé (EU Artificial Intelligence Act, 2024).

Avec le projet de loi C-27 (*Projet de loi C-27*, 2022), le Canada adopte également la stratégie d'une « mise en droit » de l'IA. Bien que des analogies puissent être établies avec le texte européen, le projet de loi canadien pose des exigences de mise en marché moins contraignantes. Surtout, le projet de loi prévoit conférer au pouvoir réglementaire la responsabilité de compléter et de clarifier le contenu de ses dispositions, ce qui ajouterait à l'encadrement canadien de l'IA une couche de complexité et d'incertitude quant à l'exécution et à l'effet en pratique de ce dernier.

Les positions étant extrêmement tranchées, l'établissement de compromis semble de plus en plus difficile à atteindre, au point de faire émerger un risque d'adoption de mesures disproportionnées ou non équilibrées.

En porte-à-faux avec ces initiatives légales se trouve par exemple le Royaume-Uni, qui, soucieux de préserver l'innovation tout en demeurant à l'affût de possibles dérives, privilégie une approche non contraignante reposant sur la définition de cinq grands principes en matière de développement et d'utilisation de l'IA – l'équité, la transparence, la sécurité, la responsabilisation et la contestabilité – ainsi que sur la capacité des entités de régulation à réglementer les systèmes d'IA dans leurs secteurs respectifs (Department for Science, Innovation and Technology of the United Kingdom, 2023).

## Qui?

Le cas britannique nous amène ainsi à un autre grand questionnement : quels sont les acteurs qui doivent intervenir dans cette reprise en main? Cette juridiction a par exemple fait le choix, pour le moment, de ne pas traiter trop spécifiquement l'IA en envisageant le numérique dans sa généralité. Plus exactement, elle a clairement décidé d'impliquer les instances existantes (comme l'Information Commissioner's Office, ICO) plutôt que de créer de nouvelles organisations dédiées à la régulation de l'IA. En tout état de cause, au-delà de la substance que l'IA requiert quant aux principes et à leur application, une réponse institutionnelle semble incontournable. Là encore, l'Union européenne semble vouloir développer une approche administrative plus structurante, avec des entités spécifiques. Le Canada, pour sa part, de façon plus « tiède », s'en remettrait au pouvoir exécutif : le projet de loi C-27 prévoit la possibilité pour le ministre de l'Innovation, des Sciences et de l'Industrie de nommer au ministère un haut fonctionnaire, soit un Commissaire à l'intelligence artificielle et aux données, pour l'appuyer dans l'administration et dans l'application de la loi relative à l'IA.

## Première caractéristique : délégations normatives

Cela dit, des traits communs peuvent être décelés dans les juridictions donnant lieu à la comparaison. Effectivement, une tendance semble se confirmer dans tous les modèles observés : les textes les plus fondateurs, idéalement les lois, réfèrent à des normes techniques qui exigent elles-mêmes des acteurs la mise en place de politiques internes afin d'objectiver leur diligence. Une double délégation s'opère alors, portée par la complexité et par l'évolution rapide du domaine. Or, il importe de prendre conscience qu'une telle manière de procéder n'est pas « neutre » et qu'elle tend en elle-même à acter une certaine forme de désengagement de l'État. Cette « agilité », que l'on semble privilégier jusque dans le corps des lois (voir le préambule du projet de loi C-27, susmentionné), n'est pas anodine et requiert une analyse critique. En premier lieu, ces normes techniques foisonnent çà et là, en grand nombre (juste pour l'Organisation internationale de normalisation (ISO), 65 normes liées à l'IA sont publiées ou sont en cours de l'être!) et dans une relative opacité, les acteurs commerciaux semblant jouir d'une certaine influence dans les forums internationaux. En deuxième lieu, une gêne est facilement identifiable dans la mesure où pour certains, ces textes ne sont pas que « techniques », mais ont une véritable portée prescriptive qui ne peut se développer dans l'entre-soi de communautés informelles dont la légitimité est mal appréhendée, notamment faute de débat démocratique. Enfin, en troisième lieu, et sans être exhaustif, l'une des problématiques les plus visibles de ce système de délégation est le caractère payant des normes. En effet, ce modèle normatif est basé sur un modèle d'affaires où ce sont les acteurs qui achètent lesdites normes pour les appliquer. Cette manière de faire, pour le moins « innovante », a notamment pour effet de favoriser les plus grosses structures et de handicaper les PMEs, moins à même de répondre à cette « agilité normative ».

## Seconde caractéristique : constat d'adolescence

Sans remettre en cause cette manière de faire, des questionnements demeurent autour de ce modèle normatif. Car les modèles juridiques, depuis l'époque moderne, sont habitués à des structures pyramidales où les flux décisionnels vont du haut vers le bas. Inversement, dans le numérique en général, et particulièrement en IA, une approche ascendante est souvent mise de l'avant. Ce changement de paradigme, s'il est sans aucun doute salvateur, n'a en revanche que peu été intégré en droit. À bien des égards, l'utilisation des différentes terminologies ayant cours (guide de conduite, lignes directrices, gouvernance, etc.) relève d'une adolescence conceptuelle qui empêche un développement serein de la manière de réguler.

# Débats scientifiques

## Internormativité

Tel que souligné précédemment, la place et la fonction du droit ainsi que son articulation avec les autres sources normatives dans la gouvernance de l'IA occupent une place centrale dans les discussions relatives à l'encadrement de cette technologie. Cette problématique soulève bien évidemment la question de l'opportunité et de la pertinence d'adopter une législation propre à l'IA, et de manière sous-jacente, celle de son degré d'agilité et de pérennité.

## Neutralité technologique

Par extension, découlent de ces enjeux les réflexions relatives à la relation entre le droit existant et la nécessité de former un nouveau corpus juridique ainsi qu'un dialogue entre le droit commun et les réglementations spécialisées. L'adoption d'un tel corpus fait actuellement débat au sein de l'écosystème québécois de l'IA ainsi qu'au sein de la communauté juridique. Plusieurs soulignent que les lois existantes, y compris les textes de droit commun et de droit spécifique du numérique, s'appliquent déjà à l'IA et pourraient faire l'objet d'adaptations pour répondre plus adéquatement aux défis spécifiques que soulèvent l'IA. Ils pointent également les risques d'une loi sur l'IA : rigidité, obsolescence, risque de fragmentation et de duplication des normes existantes ou encore atteinte au principe fondateur du droit numérique, à savoir la neutralité technologique. En marge de ces adaptations normatives qui semblent faire l'unanimité dans la communauté juridique, plusieurs arguments sont toutefois mis de l'avant pour soutenir l'adoption de textes spécifiques à l'IA : couvrir les angles morts du droit positif (*ex.* conditions de mise en marché des solutions d'IA, incluant des obligations de transparence et des dispositions en matière de gouvernance des données), imposer uniformément un développement et une utilisation responsables de l'IA au Québec et ramener la question de l'encadrement de l'IA dans la sphère du débat démocratique. Ces questions ont d'ailleurs été fréquemment soulignées dans divers rapports parus à la fin de l'automne 2023 et au début de l'hiver 2024[23].

Par extension, découlent de ces enjeux les réflexions relatives à la relation entre le droit existant et la nécessité de former un nouveau corpus juridique ainsi qu'un dialogue entre le droit commun et les réglementations spécialisées.

23 Voir entre autres : Larivière-Mantha et Hulin (2024); Gautrais et al. (2023); Marchildon et al., (2024); Beaudoin et al., (2024); Conseil de l'innovation du Québec (2024) et Autorité des marchés financiers (2024).

# Cas d'usage

## L'exemple de l'imagerie médicale

On observe dans les hôpitaux de plus en plus de systèmes d'IA (SIA) suppléer, mais plus souvent supporter, le travail des radiologues pour identifier des anomalies dans l'imagerie médicale. L'imagerie médicale constitue un exemple courant où un SIA est employé et offre des résultats probants, et ce, dans un cadre légal et institutionnel généralement très fort du fait du caractère d'intérêt public élevé.

### Cadre légal

Si l'on prend le seul cas du droit québécois, contrairement à l'impression souvent véhiculée de « vide juridique », il existe une pluralité de lois qui sont susceptibles de s'appliquer, que ce soit en matière de protection des renseignements personnels (*Loi sur la protection des renseignements personnels dans le secteur privé*. RLRQ, c. P-39.1; mieux connue sous l'appellation « loi 25 »), de données de santé (*Loi sur les renseignements de santé et de services sociaux*. RLRQ, c. R-22.1; mieux connue sous l'appellation « loi 5 »), de données biométriques (*Loi concernant le cadre juridique des technologies de l'information*. RLRQ, c. C-1.1; mieux connue sous l'abréviation « LCCJTI »), de droits fondamentaux (*Charte des droits et libertés de la personne*. RLRQ, c. C-12; mieux connue sous l'appellation « Charte québécoise ») ou encore de responsabilité (*Code civil du Québec*. RLRQ, c. CCQ-1991). Certaines comportent des dispositions assez précises et liées directement à l'IA, comme en matière de décisions automatisées par exemple. Il importe donc, en l'espèce, de savoir si la décision prise par la machine est sous le contrôle direct d'un humain et, si ce n'est pas le cas, d'en connaître les conséquences plus lourdes au niveau de la diligence et de l'information.

### Cadre principiel

Au-delà de ces textes, il est généralement demandé de s'inspirer de cadres principiels afin d'objectiver les façons de procéder. Là encore, c'est une pléthore de cadres documentaires qui sont susceptibles d'être suivis. Si plusieurs centaines de documents à saveur éthique ont vu le jour, d'autres d'une perspective juridique ont émergé tandis que des normes techniques étaient également disponibles. À bien des égards, cette diversité effraie et perturbe une communauté médicale en quête d'assurance.

### Cadre documentaire

Il en est de même sur le plan documentaire, où l'hôpital concerné doit rédiger des politiques internes explicitant sa diligence dans la collecte, dans l'utilisation et dans la communication des données. Minimalement, il se devra de déterminer la justification sociale du SIA, c'est-à-dire montrer tant le caractère probant de ces inventions, qui se basent sur des millions de clichés, que la proportionnalité associée à un tel usage de données généralement considérées comme sensibles. Également, des mesures devront être précisées et clarifiées quant à la sécurité mise en place, à la transparence offerte aux personnes concernées, à l'explicabilité du SIA et aux garanties de non-discrimination. C'est à ce prix documentaire que l'utilisation d'un SIA sera jugée responsable. Aussi, au-delà de la mise en place de tels principes, il importe de s'assurer que cet exercice puisse se faire au sein de l'institution. En l'occurrence, à part les éthiciens, les juristes et les développeurs impliqués dans la mise en place du SIA, il est considéré comme diligent de vérifier que les mesures sont adaptées auprès des médecins, des patients et des gestionnaires.

# Pistes d'action et implications pour les parties prenantes

**Sans prétendre à l'exhaustivité, il convient de souligner les éléments substantiels et institutionnels suivants.**

1

## Responsabilisation et responsabilité des acteurs de l'IA

Le domaine du numérique se caractérise par un climat de **déresponsabilisation** de ses acteurs. Remontant aux années 2000, celui-ci est alimenté par **l'incertitude** (Lacroix et Vermeys, 2023: p. 61 et s.), les **carences** (*Loi concernant le cadre juridique des technologies de l'information*. RLRQ, c. C-1.1, art. 22; voir les commentaires sur cet article dans Gautrais et al., 2023 : p. 85-87) et les **exonérations** (*Code civil du Québec*. RLRQ, c. CCQ-1991, art. 1473; Lacroix et Vermeys, 2023 : p. 104-106) qui pèsent sur l'application des régimes de responsabilité aux SIA. La *Loi sur la protection des renseignements personnels dans le secteur privé* (RLRQ, c. P-39.1) constitue un premier pas vers un renversement de cette tendance, et ce, même si son application est délicate (*Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels*. RLRQ, c. A-2.1 : art. 158 et s.; *Loi sur la protection des renseignements personnels dans le secteur privé*. RLRQ, c. P-39.1 : art. 90.1 et s.). Toutefois, les préjudices de l'IA ne se limitent pas aux atteintes à la vie privée. L'IA est susceptible de causer divers préjudices corporels (*ex.* atteinte à la vie ou à l'intégrité de la personne), matériels (*ex.* atteinte aux droits d'auteur[24]) ou moraux (*ex.* atteinte au droit à l'égalité). Le préjudice subi peut également être d'ordre individuel, collectif ou mixte. **Ce faisant, l'établissement de normes précises accompagnées de sanctions adaptées s'avère primordial pour créer un environnement où l'innovation peut prospérer sans compromettre ni la sécurité des individus ni leurs droits.**

2

## Instrumentalisation des principes juridiques existants

La recherche de solutions pour soutenir une plus grande responsabilisation des acteurs de l'IA – en l'absence d'une loi qui imposerait des obligations spécifiques en amont de la mise en marché des SIA – invite aussi à la **(re)découverte de principes juridiques existants**.

**Le principe de précaution**, qui consiste à anticiper les risques et à adopter des mesures préventives pour réduire d'éventuels effets négatifs graves et irréversibles dans un contexte d'incertitude (Centre québécois du droit de l'environnement, 2024; Botes, 2023 : p. 1), peut être appliqué à l'IA, plus particulièrement aux développeurs et aux utilisateurs.

**Les obligations fiduciaires**, telles que la loyauté et la diligence, sont également explorées comme leviers de transposition du principe de transparence dans la conception et dans le fonctionnement des algorithmes (Boine, 2023) pour renforcer la protection des renseignements personnels, pour prévenir les biais discriminatoires et, plus largement, pour intégrer de manière juridique la prise en compte des implications éthiques et sociales des applications de l'IA (Balkin, 2020; Richards et Hartzog, 2021; Benthall et Shekman, 2023). Si la réception de ces réflexions en droit québécois (Hulin, 2021; Albert-Rochette et al., 2023) doit encore faire l'objet de recherches, l'invocation du principe de précaution comme des obligations fiduciaires offre la perspective d'un subterfuge pour pallier l'adoption d'un cadre législatif spécifique à l'IA au Québec.

**Le principe de justification sociale,** rappelons-le, est un critère jurisprudentiel en droit constitutionnel canadien permettant de mesurer et de balancer les atteintes aux libertés fondamentales qui peut être instrumentalisé afin de proposer une réponse collective au développement et à l'exploitation des SIA (Gautrais et al., 2023).

24 Voir la section de l'axe Arts, médias et diversité culturelle pour les enjeux entourant le droit d'auteur.

3

## Animation normative au niveau de l'application

Au-delà des principes substantiels susmentionnés, **il est primordial de s'intéresser aussi à la réponse institutionnelle que l'on se doit d'apporter aux SIA**.

En premier lieu, si la réponse législative semble incontournable, **il est requis qu'une animation normative s'opère également au niveau de l'application**; les acteurs peinent à identifier les standards et les cadres organisationnels que la technologie exige. L'industrie, et plus particulièrement les PME, gagnerait à bénéficier d'un soutien pour l'aider à déterminer les meilleures pratiques.

En second lieu, au-delà des réflexions internationales, **il importe de rappeler qu'un débat implique une discussion transversale réunissant, au sein d'une même structure gouvernementale, des acteurs qui ne sont pas traditionnellement voués à travailler ensemble.**

4

## Participation des parties prenantes

Enfin, pour éviter les écueils susnommés liés au contexte d'urgence et de polarisation, il paraît crucial d'**adopter une approche mesurée et inclusive dans l'élaboration des réglementations portant sur l'IA.** Cela implique de prendre le temps nécessaire pour évaluer soigneusement les avantages et les inconvénients des différentes options réglementaires, pour consulter un large éventail de parties prenantes et pour s'assurer que les mesures adoptées sont à la fois efficaces pour atteindre les objectifs visés et flexibles pour s'adapter à l'évolution rapide de la technologie. Telle fut la démarche mise de l'avant par le Conseil de l'innovation du Québec dans le cadre de sa vaste campagne de consultation publique sur l'encadrement de l'IA au Québec, qui s'est déroulée au cours de l'été 2023 et dont la prise de mesures en conséquence par le gouvernement québécois se fait actuellement attendre.

# Axe
# Éthique, gouvernance et démocratie

**Sous la responsabilité de :**
Allison Marchildon et Bryn Williams-Jones

**En collaboration avec :**
Antoine Boudreau LeBlanc et Gabrielle Verreault

# Actualité scientifique

## De nombreux enjeux d'éthique et de gouvernance en IA

Si toutes les (nouvelles) technologies soulèvent des enjeux éthiques, l'arrivée de l'IA a ceci de particulier qu'elle nous a mené à porter une attention plus grande à ce type d'enjeux, et ce, tant dans les milieux académiques que sur le terrain. Nous avons ainsi vu se constituer, au cours des dernières années, une littérature scientifique foisonnante autour de l'exploration des divers enjeux éthiques soulevés par les développements dans l'intelligence artificielle (IA) (Jobin et al., 2019; Floridi et al., 2018), la robotique (Boch et al., 2021) et plus largement dans le numérique (Ashok et al., 2022). Parmi les plus discutés, mentionnons notamment les enjeux d'**accès et de protection des données** (Hulin, 2022; Auclair, 2023), de **respect et protection** de la **vie privée**[25] (Shapiro, 2010; CACM Staff, 2012) et de la **propriété intellectuelle**[26] (Lee et al., 2021; Zakir et al., 2023), d'**explicabilité** (Bélisle-Pipon et al., 2022b; Li et al., 2022; Ursin et al., 2023), de **fiabilité**, de gestion des **secrets industriels** (Marty, 2019; Meyers, 2018) ou **gouvernementaux** (Saetnan et al., 2018), de **biais** et de **discrimination** (Besse, 2020), d'**inégalités** (Conseil du statut de la femme, 2023; Marques, 2022), d'**injustices locales et internationales** (Boddington, 2023; Gentelet, 2023), d'**autonomie** des personnes, ainsi que les enjeux **environnementaux**[27], incluant l'**exploitation des ressources naturelles et énergétiques** comme les métaux stratégiques (Dauvergne, 2022; Spelda et Stritecky, 2020).

Les réflexions éthiques sur l'IA et le numérique rejoignent par ailleurs celles portant sur les enjeux de gouvernance (Larivière-Mantha et Hulin, 2024), qui sont eux aussi abondamment discutés. On s'intéresse notamment aux **pouvoirs** et **responsabilités** des **développeurs** de ces technologies, des **autorités** en charge de les implanter ou de les encadrer (Jacob et Lawarée, 2022) et du **public/société civile** (Gentelet et al., 2024; Petitgand, 2024) qui les utilise ou les subit. Dans cette foulée, la recherche de **modèles de gouvernance** adéquats afin de coordonner le développement et le déploiement de ces technologies de rupture (Boudreau LeBlanc et Williams-Jones, 2023; Bernatchez et al., 2021), de même que l'identification des **inégalités dans la distribution des pouvoirs et des responsabilités** engendrées ou renforcées par les modes de gouvernance actuels de l'IA suscitent une attention particulière.

Ces réflexions ouvrent sur les risques et potentialités d'une **gouvernance algorithmique** de nos sociétés (Langlois et Castets-Renard, 2021, Poussart, 2021) de même que sur le rôle de ces nouvelles technologies dans la transformation de la **démocratie**, de nos **institutions** et plus largement encore sur l'architecture de nos **sociétés** (Bengio, 2023; Crawford, 2021).

## Un questionnement critique et transversal

Les thématiques de l'éthique et de la gouvernance de l'IA se démarquent par leur dimension **critique** et leur **transversalité**. En effet, contrairement à d'autres thématiques abordées dans le présent état de la situation, elles ne concernent pas un domaine d'activité ou d'application précis, mais au contraire traversent l'ensemble de ceux-ci pour les questionner et les réfléchir. En tant que notion transversale, l'éthique de l'IA réfère ainsi aux réflexions **critiques** et **prospectives** entourant cette technologie émergente, et ce, peu importe ses domaines de développement et d'application et les sphères d'activité dans lesquelles elles se déploient, qu'il s'agisse de la santé, de l'éducation, du transport, du droit, du travail ou de l'environnement. Elle amène ainsi à poser un regard critique tant sur son développement (en amont) que sur son implantation et son utilisation (en aval), en interrogeant les **valeurs et principes** qui les guident, de même que leurs **impacts** sur les personnes, les communautés et l'environnement. Ce faisant, l'éthique questionne tout autant les **produits d'AI** que la **façon dont ils sont développés, commercialisés et utilisés**. Elle s'intéresse par conséquent à toute la **chaîne de valeur** de ces produits et au **cycle de vie** des données qui les alimentent, mais aussi aux **pratiques** scientifiques, organisationnelles et institutionnelles qui les entourent, en portant un regard critique sur les modes de **gestion** et de **prise de décision** entourant le déploiement de cette technologie.

Comme on peut le constater en consultant les autres sections du présent état de la situation, les enjeux d'éthique et de gouvernance de l'IA se retrouvent dans la plupart des recherches menées **au sein des différents axes de l'Obvia**, via des questionnements critiques et prospectifs, ce qui démontre bien la **transversalité** de ces deux thèmes.

25 Les enjeux entourant l'accès aux données et le respect de la vie privée sont abordés plus en détails dans les sections des axes Santé, Éducation et capacitation ainsi que Droit, cyberjustice et cybercriminalité.

26 Voir la section de l'axe Arts, médias et diversité culturelle à propos des enjeux touchant la propriété intellectuelle et le droit d'auteur.

27 Voir la section de l'axe Sobriété numérique et transition socio-écologique pour un portrait global des ces enjeux.

## Une éthique *appliquée*

Les questionnements actuels en éthique de l'IA portent par ailleurs non seulement sur des considérations et enjeux précis comme ceux précédemment évoqués, mais ils se prolongent aussi dans une réflexion sur l'action, en vue de déployer une **éthique de l'IA opératoire et appliquée**. Ce faisant, l'éthique de l'IA n'interpelle pas que les milieux académiques, mais aussi les parties prenantes de ces différents domaines d'activité, qui cherchent de plus en plus – quoiqu'à des degrés et avec des niveaux de succès variables – à intégrer une dimension éthique pratique dans le déploiement de l'IA.

Ainsi, l'éthique de l'IA participe au brouillage des frontières classiques de **l'éthique,** en fournissant un autre exemple où celle-ci ne peut plus se concevoir comme une activité théorique purement réservée à quelques **experts** spécialisés à conduire ce type de raisonnement. Plutôt, on peut parler d'une éthique **appliquée,** qui devient une **pratique concrète** et une **activité interdisciplinaire,** puisqu'elle s'appuie sur des regards et expériences multiples. À cet égard, elle mobilise souvent des **méthodes empiriques** – ex. : ethnographiques et sociologiques – ainsi que la recherche collaborative (qu'elle prenne la forme de recherche-action ou de recherche-intervention), parce qu'elles permettent une participation plus active des personnes et des communautés interpellées par l'IA (Gentelet et Bahary-Dionne, 2021; Gentelet et Mizrahi, 2024; Marchildon et al., 2024a[28]; Osmanlliu et al., 2022). Elle implique ce faisant une révision de nos modes de construction des savoirs scientifiques ainsi que ses critères de valeurs – validité, crédibilité, imputabilité, intégrité, propriété intellectuelle, etc. – qui sont d'autant plus questionnés lorsque l'essentiel du raisonnement est avancé par une machine et non plus par des humains.

Ces approches **appliquées** et **concrètes** de l'éthique se caractérisent par le fait qu'elles travaillent de plus en plus à développer des « **outils éthiques** » afin de répondre aux besoins des milieux de pratique (Morley et al., 2020; McLennan et al., 2022). Bon nombre de ces outils éthiques cherchent à **faciliter l'intégration de standards, normes, codes,** – ou toutes autres mesures techniques, politiques ou administratives – dans les pratiques des innovateurs et des entreprises, entre autres (Jobin et al., 2019; Amugongo et al., 2023). Or, devant les limites importantes des approches normatives fondées sur l'imposition et l'application de principes (Mittelstatd, 2019; Whittelestone et al., 2019), le contrôle et la conformité, d'autres approches éthiques appliquées à l'IA proposent plutôt des **outils éthiques pratiques et réflexifs** (Langlois et al., 2022; Marchildon et al., 2021; 2024a) qui cherchent pour leur part à impulser la pensée critique des développeurs, des décideurs et des utilisateurs, ainsi qu'une dynamique de dialogue entre ces différentes parties prenantes de la société et les experts. Souvent eux-mêmes co-construits avec les milieux de pratique, ces « outils éthiques » favorisent une réflexion interdisciplinaire et visent une responsabilisation et une capacitation de tous ces acteurs afin qu'ils soient outillés et mobilisés à anticiper les impacts et les différents types de risques liés à l'IA (Besse, 2020; 2021; Bengio, 2023), de même qu'à agir pour en prévenir les crises. L'éthique se croise ainsi avec la gestion des risques via des outils d'analyse des impacts concrets de l'IA, mais aussi avec ses réflexions prospectives et ses multiples propositions de visions, de valeurs et de principes pour encadrer les pratiques liées à l'IA, comme nous le verrons dans ce qui suit.

## Débats scientifiques

À l'intersection de l'éthique, de la gouvernance et de la démocratie apparaissent donc les défis de l'évaluation de l'opérationnalisation et de l'encadrement d'une IA que l'on souhaite voir déployée de façon responsable et transparente. Pour y arriver, on constate une prolifération d'activités concrètes qui se revendiquent de l'éthique. Or, si l'on s'entend assez largement sur le fait que les considérations éthiques doivent avoir une place dans le déploiement de l'IA, on retrouve toutefois derrière ces différentes approches d'importants débats scientifiques quant à la façon de mener ce déploiement: Quand et comment intégrer l'éthique (principes, analyse critique, etc.) au développement, à la commercialisation et à l'utilisation de l'IA? Comment inscrire un souci éthique dans le développement des orientations et la conception des politiques publiques, ainsi que dans les prises de décisions? Où situer les experts en éthique dans l'ensemble des décisions et opérations qui entourent l'IA? Qui sont les personnes légitimes pour « faire de l'éthique » dans ces contextes? Comment protéger le rôle critique de l'éthique en même temps qu'on suggère des solutions pratiques? Comment mobiliser l'éthique dans le déploiement de l'IA sans l'instrumentaliser (*ethics washing*)? Comment équilibrer adéquatement les bénéfices et les risques de l'IA? Comment assurer l'imputabilité de ses développeurs? Comment assurer l'alignement entre les principes éthiques énoncés et les pratiques? À quoi devrait ressembler un « audit éthique »? Quelle place donner à la population dans les réflexions autour de l'IA et son encadrement afin d'assurer un alignement avec les valeurs sociétales? Etc.

28 À noter que le *Cadre d'éthique en intelligence artificielle de l'administration publique québécoise* n'est pas encore publié en date du 1er juin 2024. Pour en savoir plus sur la démarche de conception du cadre, consulter les liens se trouvant dans la bibliographie sous Marchildon et al. (2024a).

L'objectif n'étant pas ici de faire une revue exhaustive de ces débats qui entourent l'éthique de l'IA et les diverses façons de la déployer, nous en avons plutôt retenu deux qui, selon notre expérience et une revue des recherches des membres de notre axe, nous ont semblé particulièrement prégnants au cours des deux dernières années.

## La multiplication des déclarations et cadres éthiques de l'IA

En quelques années à peine, les initiatives visant à assurer un développement de l'IA qui soit « éthique», « responsable » ou « digne de confiance » se sont multipliées, donnant lieu à une panoplie de déclarations de principes, chartes éthiques ou approches de conception éthique (*ethics by design*). Ces projets et d'autres initiatives émanent tout autant de développeurs d'IA et joueurs de l'industrie que d'États, associations professionnelles, organisations internationales ou de la société civile.

Dans la littérature scientifique, les chercheuses et chercheurs se sont rapidement intéressés au phénomène, d'abord pour le décrire et en identifier les convergences et divergences (Jobin et al., 2019; Fjeld et al, 2020; Tidjon et Khomh, 2022), mais aussi pour débattre de la valeur et de l'impact réel de ces principes de « haut niveau » sur les orientations en matière d'IA. Car bien que les principes mis de l'avant dans ces cadres constituent souvent le point de départ d'une réflexion sur les enjeux soulevés par l'IA et qu'ils demeurent par conséquent pour plusieurs un passage obligé en éthique, les critiques à l'égard des nombreuses déclarations de principes et chartes éthiques en matière d'IA sont elles aussi de plus en plus nombreuses. Ainsi, plusieurs chercheurs soutiennent que les principes ne peuvent, à eux seuls, garantir une « IA éthique » (Mittelstadt, 2019) et que sa mise en œuvre doit aussi s'appuyer sur un ensemble d'autres mesures et institutions. Or, à l'heure actuelle, on considère que cette opérationnalisation est trop souvent floue (Maclure et Saint-Pierre, 2018) et que seule une faible proportion de ces chartes incluent des recommandations ou des exemples pour leur mise en œuvre (AlgorithmWatch, 2020). En outre, les disparités entre les principes énoncés dans ces chartes, leur nombre élevé et le fait qu'ils sont souvent déconnectés de la réalité pratique fait en sorte qu'ils deviennent ni plus ni moins des « listes d'épicerie » (Martin et al., en rédaction) qui créent plus de confusion qu'ils ne guident les acteurs sur le terrain.

En parallèle, d'autres critiques dénoncent le fait que, malgré les nombreux principes proposés pour guider le développement de l'IA, certains demeurent notablement manquants. C'est notamment le cas des principes en lien avec les enjeux environnementaux, ou affirmant l'importance de la solidarité et de la dimension relationnelle de la technologie et de l'IA, qui gagnerait à être développée. Souvent formulées par des actrices et acteurs ou chercheuses et chercheurs s'inscrivant dans des courants environnementaux, féministes, ou des études critiques et culturelles, ces critiques appellent à une « décolonisation » de l'IA et de l'éthique de l'IA (Adams, 2021; Dignum, 2022; Sabelo, 2020; Mohamed, Png et Isaac, 2020) afin qu'elle reflète mieux les préoccupations et soucis réels de tou.te.s les personnes qui vivront les effets de l'IA, et non seulement ceux de ses concepteurs ou des états qui ont les moyens de l'encadrer.

Bref, les débats autour des déclarations de principes et chartes éthiques en matière d'IA sont nombreux, et entraînent dans la foulée plusieurs autres questions, notamment à savoir comment mettre de l'avant une éthique qui ne soit pas que discours et promesses, ou encore comment s'assurer que les principes qui sont mis de l'avant ne seront pas le reflet que d'une fraction de la population.

## Le *ethics bashing*

Au cours des dernières années, une critique de plus en plus courante se fait entendre à l'effet que l'éthique serait un frein à l'innovation, critique souvent associée à ce que l'on pourrait appeler du "*ethics bashing*" (Bietti, 2020) et que l'on pourrait traduire par « dénigrement de l'éthique ». Ceux qui la formulent avancent que ce serait le cas notamment parce que l'éthique appartient au registre de la déclaration ou de la bonne volonté et qu'elle n'a ce faisant « pas de dents ». Or, une telle critique semble traduire une confusion entre éthique et droit qu'il importe de clarifier. En effet, alors que le droit a une fonction explicitement normative et s'appuie sur des pouvoirs de coercition pour forcer certains comportements et en empêcher d'autres, parfois par la sanction, l'éthique peut être considérée comme plus « douce » parce qu'elle n'a pas cette fonction de coercition. Mais son pouvoir est d'une autre nature : il consiste plutôt à influencer les comportements par, entre autres, l'invitation à mener une réflexion critique sur l'IA, à clarifier et justifier certaines décisions, à promouvoir l'identification et la gestion des enjeux et risques liés à l'IA comme les biais algorithmiques, ou à trouver des points de consensus où les valeurs et les intérêts des différentes parties prenantes s'alignent. De plus, l'éthique tire aussi son pouvoir et sa pertinence du fait qu'elle utilise un langage accessible et significatif pour un large éventail de parties prenantes : parler de justice, d'équité, de responsabilité, de valeur, etc. sont des idées qui ont du sens pour bien des personnes. L'éthique permet ainsi de nommer les problèmes (**identifier**), de comprendre leur nature et leurs implications (**évaluer**) et implique souvent le devoir que quelqu'un trouve une solution acceptable pour tous (**résoudre/de gérer**).

Conséquemment, tant les critiques consistant à dénigrer les initiatives éthiques en matière d'IA que les réponses qui y sont adressées mettent en lumière que la contribution de l'éthique dans le domaine de l'IA réside dans sa problématisation nuancée de questions complexes et transversales, une contribution qui s'avère fort différente de celle du droit, mais néanmoins essentielle. Ainsi, dans une visée de promouvoir un meilleur encadrement de l'IA, un nombre croissant de chercheuses et chercheurs avancent que le recours à l'éthique gagnerait à être non pas dénigré, mais plutôt vu comme complémentaire au droit, afin de combiner leurs forces respectives (Bietti, 2020; Seger, 2022; Koniakou, 2023).

> Mais son pouvoir [celui de l'éthique] est d'une autre nature : il consiste plutôt à influencer les comportements par, entre autres, l'invitation à mener une réflexion critique sur l'IA, à clarifier et justifier certaines décisions, à promouvoir l'identification et la gestion des enjeux et risques liés à l'IA comme les biais algorithmiques, ou à trouver des points de consensus où les valeurs et les intérêts des différentes parties prenantes s'alignent.

# Cas d'usage

## ChatGPT, l'IA générative et l'éthique

Fin 2022, les débats sur la responsabilité et la confiance en matière d'IA ont quitté la sphère spéculative pour prendre forme dans la réalité avec l'arrivée soudaine de ChatGPT, le modèle d'IA générative commercialisé par OpenAI. Comme l'ont souligné des membres de l'Obvia dans un document de réflexion sur les enjeux de l'IA générative (Marchildon et al., 2024b), l'arrivée de cette technologie disruptive soulève de nombreux enjeux éthiques et de gouvernance, et ce, à plusieurs niveaux : sur la technologie elle-même et sa conception; dans les différents domaines où elle est utilisée; dans la façon dont les entreprises la commercialisent; dans l'encadrement légal qu'on en fait (ou que l'on ne fait pas). Parmi les principaux enjeux que soulève l'IA générative, mentionnons notamment ceux liés à la propriété intellectuelle, aux biais - avec les risques de discrimination et les inégalités que cela entraîne ou renforce – , à la manipulation d'informations ou encore aux risques pour la démocratie et aux impacts environnementaux (Marchildon et al., 2024b). Son impact sur l'éducation et en ce qui a trait au monde du travail s'avère particulièrement majeur, puisque l'IA générative y opère des changements fondamentaux dans nos modes de fonctionnement, dans notre rapport au monde, dans notre accès à l'information et au savoir ainsi que dans nos modes d'interaction qui engendreront dans leur foulée nombre d'autres enjeux (Marchildon et al., 2024b).

Photo : Diego - stock.adobe.com

Or, à l'heure actuelle, ni les entreprises qui commercialisent ces technologies, ni les institutions (ex. : éducation, fonction publique), ni les États ne se sont montrés en mesure de prévenir de façon adéquate les enjeux qui découlent de son utilisation dans ces différentes sphères. En effet, la réglementation sur l'IA, encore à l'état de projet dans la plupart des pays, ne couvrait tout simplement pas l'IA générative au moment où ChatGPT est arrivé, faisant en sorte que les projets de loi et les politiques en préparation ont dû être adaptés in extremis. En ce qui concerne les entreprises impliquées dans la commercialisation de ChatGPT, soit OpenAI et Microsoft, leur prise de responsabilité s'est avérée essentiellement réactive et elles ont eu tendance à instrumentaliser le discours éthique, en poursuivant plutôt une stratégie agressive axée sur les gains de parts de marché.

# Pistes d'action et implications pour les parties prenantes

**À la lumière de ce qui précède, on constate que la réflexion éthique appliquée à l'IA est à la fois présente et nécessaire à tous les stades de son développement, dans sa gouvernance et dans tous ses champs d'application. Mais dans la même foulée, on peut aussi constater que la façon de mener cette réflexion et d'intégrer les considérations éthiques dans le déploiement de l'IA et dans sa gouvernance est par ailleurs sujette à de nombreux débats et mécompréhensions.**

**Les membres de l'axe Éthique, gouvernance et démocratie de l'Obvia travaillent pour leur part au déploiement d'approches constructives et capacitantes de l'éthique, dont le rôle ne saurait se réduire à la seule énonciation de principes, ainsi qu'à la construction d'une gouvernance axée sur la collaboration et la participation des différents acteurs touchés par l'IA. Afin de poursuivre le développement de cette perspective, les pistes d'action suivantes sont à privilégier :**

1

Favoriser la **réflexion éthique et critique des différentes parties prenantes touchées par l'IA, incluant la société civile et les décideurs publics**, et ce, par la littératie en IA, la sensibilisation et l'éducation en matière d'éthique, mais aussi en développant des outils et mesures organisationnelles, institutionnelles et organisationnelles qui leur permettront de déployer une telle réflexion.

2

Favoriser la **participation des différentes parties prenantes touchées par l'IA, incluant la société civile, dans la gouvernance de l'IA** et mettre en place des mécanismes organisationnels, institutionnels et organisationnels qui leur permettront une telle participation.

3

Intégrer les **considérations et réflexions éthiques au cœur de la gouvernance de l'IA**, afin que celles-ci soient présentes dans toutes les décisions touchant l'IA, depuis l'amont de son développement jusqu'à son aval.

4

Mener des **recherches collaboratives impliquant chercheur.e.s et développeurs d'IA, de même que des décideurs publics** afin de développer des outils et mesures éthiques, ainsi que des politiques publiques et approches de gouvernance ancrés à la fois dans les besoins réels du terrain et en phase avec les plus récentes connaissances en matière d'éthique.

# Conclusion

S'inscrivant dans la perspective interdisciplinaire et intersectorielle au cœur même de la mission de l'Obvia, le présent document a permis de réunir les expertises des 7 axes de recherche thématiques afin de mettre en lumière l'état actuel des connaissances ainsi que les principaux enjeux à propos des impacts sociétaux de l'IA et du numérique. De cette mise en commun a émergé un nombre important de risques, d'opportunités et de débats, mais également de pistes d'actions à envisager pour les différentes parties prenantes. Si plusieurs enjeux sont spécifiques aux différents axes et disciplines, il a également été possible de constater que plusieurs d'entre eux sont communs à certains axes, voire transversaux. Alors que le déploiement à grande échelle des modèles d'IA générative a touché l'ensemble des sphères de la société, plusieurs enjeux soulevés par ces modèles sont liés à leur fonctionnement même. À titre d'exemple, les questions relatives aux données utilisées pour entraîner et faire fonctionner ces modèles touchent l'ensemble des domaines et regroupent à la fois les enjeux relevant du respect de la vie privée, des renseignements personnels, de la propriété intellectuelle et du droit d'auteur, notamment soulignés par les axes Santé durable, Éducation et capacitation, Éthique, gouvernance et démocratie et Droit, cyberjustice et cybersécurité. De même, la consommation énergétique et les coûts environnementaux engendrés par les modèles d'IA générative, détaillés dans l'axe Sobriété numérique et transition socio-écologique, sont des enjeux inhérents à l'utilisation de cette technologie et concernent à cet égard l'ensemble des axes et des secteurs d'activité.

En ce qui a trait aux différents débats scientifiques soulevés par les axes, au-delà de leurs spécificités respectives, il a été possible de faire ressortir des questionnements à propos de la valeur même de certaines utilisations de l'IA. En effet, bien que plusieurs avantages des SIA ne soient plus à démontrer, notamment dans une logique d'efficience et de performance, ces avantages ne se traduisent pas forcément au niveau du bien commun ou du bien-être des individus et vont même parfois à l'encontre de valeurs et de principes fondamentaux. Sans prétendre offrir des réponses définitives à ces débats complexes, le fait d'interroger et de remettre en question de manière critique l'implantation de cette technologie dans nos sociétés, trop souvent considérée comme inéluctable et relevant du progrès, constitue en soi un exercice essentiel.

En terminant, plusieurs pistes d'actions mises de l'avant par les axes et destinées aux différentes parties prenantes regroupent des éléments communs. Ces pistes d'actions visent : l'encadrement (éthique ou juridique) de l'IA, l'intégration des considérations éthiques et sociétales dans le développement de l'IA, le développement de la recherche et des connaissances sur les impacts de l'IA, l'amélioration de la littératie numérique et des compétences en IA ainsi que l'implication et la participation de l'ensemble des parties prenantes, incluant la société civile, autant au processus de développement de l'IA qu'à son encadrement. Il est ici nécessaire de souligner comment ces recommandations sont interreliées et vont, en ce sens, de pair. Ainsi, l'implication des différentes parties prenantes et du grand public nécessite un certain niveau de littératie numérique et de compréhension de l'IA et de ses enjeux. De la même manière, la prise en compte d'enjeux éthiques et sociétaux dans l'ensemble du développement de l'IA et dans son encadrement repose sur des connaissances suffisantes et à jour des impacts sociétaux de l'IA et du numérique. Continuer à documenter ces impacts demeure donc primordial, mais il s'avère également nécessaire de développer la recherche sur la manière d'opérationnaliser et de mettre en place les différentes solutions envisagées pour s'assurer que l'implantation de l'IA dans nos sociétés se fasse bel et bien aux bénéfices de toutes et tous. Le présent état de la situation s'inscrit justement dans cette volonté et témoigne de l'engagement soutenu de l'Obvia envers le développement de la recherche et des connaissances sur les impacts sociétaux de l'IA et du numérique.

# Bibliographie

## Introduction


Arendt, H. (1958). *The Human Condition*. University of Chicago Press.

Bengio, Y., Lequesne, C., Loiseau, H., Maclure, J., Powell, J., Solomun, S. et Langlois, L. (2024). *Dialogues interdisciplinaires : les risques majeurs de l'IA générative*. Obvia, IVADO, Mila. https://doi.org/10.61737/YRCP3187

Bourgeot, R. (2023). *Sommet de l'IA de Bletchley Park : Concertation mondiale ou lobbying chic ?*. IRIS-Institut de Relations Internationales et Stratégiques. https://www.iris-france.org/179597-sommet-de-lia-de-bletchley-park-concertation-mondiale-ou-lobbying-chic/

Conseil de l'innovation du Québec. (2023). *Investissements publics favorisant le développement et le déploiement responsables de l'IA auprès du secteur privé: État de situation* https://conseilinnovation.quebec/wp-content/uploads/2023/10/CIQ_Investissements_IA_secteur_prive_EDS-1.pdf

Conseil de l'innovation du Québec. (2024). *Prêt pour l'IA – Répondre au défi du développement et du déploiement responsables de l'IA au Québec*. https://conseilinnovation.quebec/wp-content/uploads/2024/02/Rapport_IA_CIQ-1.pdf

Future of Life Institute. (2023). *Pause Giant AI Experiments: An Open Letter*. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Khoury, R. (2017). Combattre les comportements toxiques sur Internet avec l'intelligence artificielle [entrevue et article de Jonathan Lavoie]. *Radio-Canada*. https://ici.radio-canada.ca/nouvelle/1064562/internet-lintelligence-artificielle-radicalisation-intimidation--harcelement-sexuel-universite-laval

MacAskill, W. (2022). *What We Owe the Future*. Basic Books.

PwC. (2021). *Analyse économique des investissements réalisés en intelligence artificielle au Québec* (produit pour le compte du forum IA Québec). https://api.forum-ia.devbeet.com/app/uploads/2022/03/pwc_forumiaqc_sommaire_2022.pdf


## Axe Santé durable


Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Alderem, A., Alrashed, M., Bin Saleh, K., Badreldin, H.A., Al Yami, M. J., Al Harbi, S. et Albekairy, A. M. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*, *23*(1), 689. https://doi.org/10.1186/s12909-023-04698-z

Awasthi, R., Mishra, S., Grasfield, R., Maslinski, J., Mahapatra, D., Cywinski, J. B., Khanna, A.K. et al. (2023). Artificial Intelligence in Healthcare: 2023 Year in Review. *medRxiv*. , https://doi.org/10.1101/2024.02.28.24303482.

Balcombe, L. (2023). AI Chatbots in Digital Mental Health. *Informatics*, 10(4), 82. https://doi.org/10.3390/informatics10040082.

Barocas, S. et Nissenbaum, H. (2014). Big data's end run around anonymity and consent. *Privacy, big data, and the public good: Frameworks for engagement*. https://doi.org/10.1017/CBO9781107590205.004.

Bender, E.M., Gebru, T., McMillan-Major, A. et Shmitchell, S. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Dans *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021. https://doi.org/10.1145/3442188.3445922.

Benaich, N. (2023). *State of AI Report 2023*. Air Street Capital. Consulté le 24 mars 2024. https://www.stateof.ai/.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein M. S. et al. (2021). On the Opportunities and Risks of Foundation Models ». *arXiv preprint*. https://doi.org/10.48550/arXiv.2108.07258.

Boverhof, B. J., Redekop, W. K., Bos, D., Starmans, M. P., Birch, J., Rockall, A. et Visser, J. J. (2024). Radiology AI Deployment and Assessment Rubric (RADAR) to bring value-based AI into radiological practice. *Insights into Imaging*, *15*(1), 34. https://doi.org/10.1186/s13244-023-01599-z.

Brahimi, L. et Legendre, M.-C. (2023). *Évaluation et implantation des solutions numériques en santé : regard sur la participation des usagers*. INESSS, Institut national d'excellence en santé et services sociaux. Consulté le 23 août 2023. https://www.inesss.qc.ca/fileadmin/doc/INESSS/Rapports/OrganisationsSoins/INESSS_Partenariat_telesante_EC.pdf.

Clusmann, J., Kolbinger, F. R., Muti, H. S., Caxrrero, Z. I., Eckardt, J. N., Laleh, N. G., Löffler, C.M.L., Schwarzkopf, S.-C., Unger, M., Veldhuizen, G. P., Wagneret, S.J. et Kather, J. N. (2023). The future landscape of large language models in medicine. *Communications medicine*, *3*(1), 141. https://doi.org/10.1038/s43856-023-00370-1.

Comité permanent de l'industrie et de la technologie. (2023). Correspondance de l'honorable François-Philippe Champagne, ministre de l'Innovation, des Sciences et de l'Industrie - Amendements à la LIAD. https://www.ourcommons.ca/content/Committee/441/INDU/WebDoc/WD12751351/12751351/MinisterOfInnovationScienceAndIndustry-2023-11-28-Combined-f.pdf

Commission de la santé mentale du Canada. (2021). L'intelligence artificielle dans les services de santé mentale : une analyse de l'environnement. https://mentalhealthcommission.ca/wp-content/uploads/drupal/2021-08/AI_enviorn_scan_brief_fr.pdf.

Cumyn, A., Barton, A., Dault, R., Cloutier, A. M., Jalbert, R. et Ethier, J. F. (2020). Informed consent within a learning health system: A scoping review. *Learning health systems*, *4*(2), e10206. https://doi.org/10.1002/lrh2.10206.

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Schieber, S.,Waldo, J., Weinberger, D., Weller, A. et Wood, A. (2017). Accountability of AI under the law: The role of explanation. *arXiv preprint* http://arxiv.org/abs/1711.01134.

Europe. (2024). *Règlement sur l'intelligence artificielle*. Les textes de loi – Projet final. https://artificialintelligenceact.eu/fr/l-acte/

Fournier-Tombs, E et McHardy, J. (2023). A medical ethics framework for conversational artificial intelligence. *Journal of Medical Internet Research*, *25*, e43068. https://doi.org/10.2196/43068.

Froomkin, A. Michael. (2019). Big Data: Destroyer of Informed Consent. *Yale Journal of Law and Technology*. https://papers.ssrn.com/abstract=3405482.

Hughes, D. (2023). *La transformation numérique du réseau de la santé et des services sociaux en vue d'intégrer l'intelligence artificielle : un regard éthique*. Commission de l'éthique en science et technologie (CEST), Gouvernement du Québec. https://www.ethique.gouv.qc.ca/media/3kcdmpkp/cest_transformation-numerique-rsss-ai_epac.pdf.

Hulin, A.-S. (2022). La gouvernance des données en droit civil québécois : comment (re)concilier protection et exploitation des données personnelles ?. *Revue internationale de droit économique* t.XXXVI, no 3, 39-61. https://doi.org/10.3917/ride.363.0039.

Global Digital Health Monitor. (2023) The State of Digital Health 2023. Global Development Incubator, https://static1.squarespace.com/static/5ace2d0c5cfd792078a05e5f/t/656f97969301e337ada15270/1701812128734/State+of+Digital+Health_2023.pdf.

Joshi, G., Jain, A., Araveeti, S. R., Adhikari, S., Garg, H., et Bhandari, M. (2024). FDA-Approved Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices: An Updated Landscape. *Electronics*, *13*(3), 498. https://doi.org/10.3390/electronics13030498.

Kak, A. (2023). Make No Mistake—AI Is Owned by Big Tech. *MIT Technology Review*. https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech/.

Keierleber, M. (2022)Young and Depressed? Try Woebot! The Rise of Mental Health Chatbots in the US. *The Guardian*, 13 avril 2022, sect. US news. https://www.theguardian.com/us-news/2022/apr/13/chatbots-robot-therapists-youth-mental-health-crisis.

Khawaja, Z. et Bélisle-Pipon, J. C. (2023). Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. *Frontiers in Digital Health*, *5*, 1278186. https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2023.1278186.

Lavoie-Moore, M. (2023). Portrait de l'intelligence artificielle en santé au Québec. Propositions pour un modèle d'innovation au profit des services et des soins de santé publics. IRIS. https://iris-recherche.qc.ca/publications/intelligence-artificielle-sante/.

Van Leeuwen, K. G., de Rooij, M., Schalekamp, S., van Ginneken, B. et Rutten, M. J. (2022). How does artificial intelligence in radiology improve efficiency and health outcomes?. *Pediatric Radiology*, 1-7.https://doi.org/10.1007/s00247-021-05114-8.

*Loi sur les renseignements de santé et de services sociaux et modifiant diverses dispositions législatives*. (2023) LQ 2023, c 5. https://canlii.ca/t/6f6mg.

Ma, B., Yang, J., Wong, F. K. Y., Wong, A. K. C., Ma, T., Meng, Zhao, Y., Wang, Y. et Lu, Q. (2023). Artificial intelligence in elderly healthcare: A scoping review. *Ageing Research Reviews*, 83. https://doi.org/10.1016/j.arr.2022.101808.

Medium. (2022, 7 décembre). The Potential of Large Language Models(LLMs) in Healthcare: Improving Quality of Care and Patient outcomes. *Medium* (blog). https://medium.com/@BuildGP/the-potential-of-large-language-models-in-healthcare-improving-quality-of-care-and-patient-6e8b6262d5ca.

Moodley, K., & Rennie, S. (2023). ChatGPT has many uses. Experts explore what this means for healthcare and medical research. *The Conversation*. http://theconversation.com/chatgpt-has-many-uses-experts-explore-what-this-means-for-healthcare-and-medical-research-200283.

Nicol, E. D., Weir-McCall, J. R., Shaw, L. J. et Williamson, E. (2023). Great debates in cardiac computed tomography: OPINION:"artificial intelligence and the future of cardiovascular CT–Managing expectation and challenging hype". *Journal of Cardiovascular Computed Tomography*, *17*(1), 11-17. https://doi.org/10.1016/j.jcct.2022.07.005.

Marchildon, A., Boine, C., Sabourin Laflamme, A., Anctil, D., Boudreau LeBlanc, A., Auclair, S., Balagué, C., Morin-Bertrand, F.-A., Castets-Renard, C., Jackson, P. et Langlois, L. (2024). *Un an après l'arrivée de ChatGPT : Réflexions de l'Obvia sur les enjeux et pistes d'action possibles face à l'IA générative*. Obvia. https://doi.org/10.61737/ZKWZ3721

OCDE. *L'IA dans le domaine de la santé - Un énorme potentiel, d'énormes risques*. https://www.oecd.org/fr/sante/IA-en-sante-immense-potentiel-enormes-risques.pdf.

Omarov, B., Narynov, S. et Zhumanov, Z. (2023). Artificial Intelligence-Enabled Chatbots in Mental Health: A Systematic Review. *Computers, Materials & Continua*, *74*(3). https://doi.org/10.32604/cmc.2023.034655.

Organisation mondiale de la Santé (OMS). (2024). *Ethics and governance of artificial intelligence for health. Guidance on large multi-modal models*. https://www.who.int/fr/news/item/18-01-2024-who-releases-ai-ethics-and-governance-guidance-for-large-multi-modal-models.

Panch, T., Mattie, H. et Celi, L. A. (2019). The "inconvenient truth" about AI in healthcare. *NPJ digital medicine*, *2*(1), 1-3. https://doi.org/10.1038/s41746-019-0155-4.

Pégny, M. et Ibnouhsein, I. (2018). Quelle transparence pour les algorithmes d'apprentissage machine ? *Revue d'intelligence artificielle* 32: 447. https://hal.science/hal-01877760/.

Ploug, T. et Holm, S. (2016). Meta consent–a flexible solution to the problem of secondary use of health data. *Bioethics*, *30*(9), 721-732. *Bioethics* 30, no 9: 721-32. https://doi.org/10.1111/bioe.12286.

Projet de loi C-27. (2022). *Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois* (présenté en première lecture le 16 juin 2022 à la Chambre des Communes du Canada).

Qiu, J., Li, L., Sun, J., Peng, J., Shi, P., Zhang, R., ... et Lo, B. (2023). Large ai models in health informatics: Applications, challenges, and the future. *IEEE Journal of Biomedical and Health Informatics*.https://doi.org/10.1109/JBHI.2023.3316750.

Rani, S., Kataria, A., Kumar, S. et Tiwari, P. (2023). Federated learning for secure IoMT-applications in smart healthcare systems: A comprehensive review. *Knowledge-Based Systems*, 110658. https://doi.org/10.1016/j.knosys.2023.110658.

Ribeiro, M., Sameer Singh, T. et Guestrin, C. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier ». *arXiv*. http://arxiv.org/abs/1602.04938.

Rioux, M., O'Donnell, V. et Alaoui, S. (2024). *Guide pratique en technologies éducatives: Vers une conception judicieuse et responsable pour les milieux éducatifs*. Edteq, CTREQ, Obvia. Québec. https://www.obvia.ca/ressources/guide-pratique-en-technologies-educatives-vers-une-conception-judicieuse-et-responsable-pour-les-milieux-educatifs

Roppelt, J. S., Kanbach, D. K., & Kraus, S. (2023). Artificial intelligence in healthcare institutions: A systematic literature review on influencing factors. *Technology in society*, 102443. https://doi.org/10.1016/j.techsoc.2023.102443.

Santé Canada. (2023). *Ébauche des lignes directrices préalables à la mise en marché pour les instruments médicaux fondés sur l'apprentissage machine*. Santé Canada. https://www.canada.ca/fr/sante-canada/services/medicaments-produits-sante/instruments-medicaux/information-demandes/lignes-directrices/prealables-mise-marche-instruments-medicaux-fondes-apprentissage-machine.html.

Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham L. Rev.*, *87*, 1085. https://heinonline.org/HOL/LandingPage?handle=hein.journals/flr87&div=44&id=&page=

Silcox, C., Zimlichmann, E., Huber, K., Rowen, N., Saunders, R., McClellan, M., ... et Bates, D. W. (2024). The potential for artificial intelligence to transform healthcare: perspectives from international health leaders. *NPJ Digital Medicine*, *7*(1), 88. https://doi.org/10.1038/s41746-024-01097-6.

Stanford University. (2024). *AI Index Report 2023 – Artificial Intelligence Index*. The Stanford Institute for Human-Centered Artificial Intelligence (HAI). https://aiindex.stanford.edu/report/.

SciencesPo-Chaire Santé. (2023). *Enjeux de l'intelligence artificielle en santé*.https://www.sciencespo.fr/chaire-sante/sites/sciencespo.fr.chaire-sante/files/Enjeux%20de%20l%27IA%20en%20sante%CC%81%2026.05.pdf

Tornero-Costa, R., Martinez-Millana, A., Azzopardi-Muscat, N., Lazeri, L., Traver, V., et Novillo-Ortiz, D. (2023). Methodological and quality flaws in the use of artificial intelligence in mental health research: systematic review. *JMIR Mental Health*, *10*(1), e42045. https://doi.org/10.2196/42045.

UNESCO. (2017). *Rapport du CIB sur les mégadonnées et la santé*. https://unesdoc.unesco.org/ark:/48223/pf0000248724_fre.locale=fr.

Vilaza, G. N., et McCashin, D. (2021). Is the automation of digital mental health ethical? Applying an ethical framework to chatbots for cognitive behaviour therapy. *Frontiers in Digital Health*, *3*, 689736. https://doi.org/10.3389/fdgth.2021.689736.

Wang, J., Liu, Z., Zhao, L., Wu, Z., Ma, C., Yu, S., ... et Zhang, S. (2023). Review of large vision models and visual prompt engineering. *Meta-Radiology*, 100047. https://doi.org/10.1016/j.metrad.2023.100047.

Woebot Health. (s.d.). Woebot Health. https://woebothealth.com/.

Zahlan, A., Ranjan, R. P. et Hayes, D. (2023). Artificial intelligence innovation in healthcare: Literature review, exploratory analysis, and future research. *Technology in society*, 102321.https://doi.org/10.1016/j.techsoc.2023.102321.

Zhang, P., & Kamel Boulos, M. N. (2023). Generative AI in medicine and healthcare: Promises, opportunities and challenges. *Future Internet*, *15*(9), 286. https://doi.org/10.3390/fi15090286.

Zhuravel, H. (2023). How IoMT is Impacting Healthcare Delivery. *Binariks* . Consulté le 21 mars 2024. https://binariks.com/blog/internet-of-medical-things-iomt/.

## Axe Éducation et capacitation


Arseneau, R., Labbé, M., Brown, H. et Voyer, S. (2024). *Guide pratique en technologies éducatives : vers une intégration judicieuse et responsable en milieu éducatif*. Edteq, CTREQ, Obvia. Québec. https://www.obvia.ca/ressources/guide-pratique-en-technologies-educatives-vers-une-integration-judicieuse-et-responsable-en-milieu-educatif

Bléhaut, M., Ierget, J., Serreau, M. et Plantard, P. (2021). *La société numérique française : Définir et mesurer l'éloignement numérique*. République française. https://agence-cohesion-territoires.gouv.fr/sites/default/files/2023-05/PUBLICATION%20ANCT_Soci%C3%A9t%C3%A9%20num%C3%A9rique_CREDOC_CREAD_2023_RAPPORT-La%20soci%C3%A9t%C3%A9%20num%C3%A9rique%20fran%C3%A7aise%20-%20d%C3%A9finir%20et%20mesurer%20l%27%C3%A9loignement%20du%20num%C3%A9rique.pdf

Celik, I., Dindar, M., Muukkonen, H., et Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, 66(4), 616-630.

Chen, M., Cheng, H., Du, Y., Xu, M., Jiang, W., et Wang, C. (2023, June). Two wrongs don't make a right: Combating confirmation bias in learning with label noise. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 12, pp. 14765-14773).

Collard, A. M. (2024, 12 février). *4 ways to future-proof against Deepfakes in 2024 and beyond*. World Economic Forum. https://www.weforum.org/agenda/2024/02/4-ways-to-future-proof-against-deepfakes-in-2024-and-beyond/

Collin, S., Denouël, J., Guichon, N., Schneider, E. (2022). *Le numérique en éducation et formation. Approches critiques*. Paris, Presses des Mines.

Collin, S., Guichon, N. et Ntébutsé, J.-G. (2015). Une approche sociocritique des usages numériques en éducation. STICEF, 22, 89-117.

Collin, S., Lacelle, N., Ladouceur, M.A., Nicole, M.C., Petit, B. et Bourgeois, C. (2024). Derrière les écrans : acteurs, usages, enjeux et contextes du numérique en éducation (Sommet du numérique, Panel/table ronde). https://2024.sommetnumerique.ca/fr/contributions/details/838

Collin, S., Lepage, A., et Nebel, L. (2023). Enjeux éthiques et critiques de l'intelligence artificielle en éducation: une revue systématique de la littérature. *Canadian Journal of Learning and Technology*, 49(4), 1-29. https://www.erudit.org/en/journals/cjlt/2023-v49-n4-cjlt09182/1110083ar/abstract/

Collin, S., et Marceau, E. (2022). Enjeux éthiques et critiques de l'intelligence artificielle en enseignement supérieur. *Éthique publique. Revue internationale d'éthique sociétale et gouvernementale, vol. 24, n° 2*, Article 24, n° 2. https://doi.org/10.4000/ethiquepublique.7619

Conseil de l'innovation du Québec. (2024). Prêt pour l'IA. https://conseilinnovation.quebec/wp-content/uploads/2024/02/Rapport_IA_CIQ-1.pdf

Conseil supérieur de l'éducation (CSE) et Commission de l'éthique en science et en technologie (CEST). (2024). Intelligence artificielle générative en enseignement supérieur : enjeux pédagogiques et éthique. https://www.cse.gouv.qc.ca/wp-content/uploads/2024/04/50-0566-SO-IA-generative-enseignement-superieur-enjeux-ethiques.pdf

Courrier international. (2024, 22 mars). Éducation : Après avoir misé sur le numérique à l'école, les pays scandinaves font machine arrière. https://www.courrierinternational.com/article/education-apres-avoir-mise-sur-le-numerique-a-l-ecole-les-pays-scandinaves-font-machine-arriere

Coalition for Content Provenance and Authenticity. (s.d.). https://c2pa.org/

Davidson, A.-L.et Charbonneau, O. (2024). Comment entrevoir l'enseignement universitaire à l'ère de l'IA générative. *In* Stéphane Allaire, Descheneaux, F. Université du futur. Chapitre 8, pp. 111-125. Presses de l'université du Québec.

Denouël, J. (2023). Transformations pédagogiques numériques dans l'enseignement supérieur : une approche sociologique critique, *Vol.1, Note de synthèse, Habilitation à diriger des recherches*, Université Rennes 2

Deschamps, T. (2024, 10 janvier). Les compétences en intelligence artificielle sont prisées par les employeurs. *La Presse*. https://www.lapresse.ca/affaires/techno/2024-01-10/les-competences-en-intelligence-artificielle-sont-prisees-par-les-employeurs.php

Dion-Viens, D. (2024). France: des recommandations choc pour limiter l'usage des écrans chez les jeunes. *Le Journal de Québec*. https://www.journaldequebec.com/2024/05/04/france--des-recommandations-choc-pour-limiter-lusage-des-ecrans-chez-les-jeunes

Fitzpatrick, C., Lemieux, A., Smith, J., West, G. L., Bohbot, V. et Asbridge, M. (2023). Is adolescent internet use a risk factor for the development of depression symptoms or vice-versa? *Psychological Medicine, 53*(14), 6773–6779. https://doi.org/10.1017/S0033291723000284

Gautherie, A. (2024). Elections : 20 géants de la tech s'allient pour lutter contre les deepfakes. *Journal du Geek*. https://www.journaldugeek.com/2024/02/19/elections-20-geants-de-la-tech-sallient-pour-lutter-contre-les-deepfakes/

Giroux, P., Dumouchel, G., Hébert, É. et Agbotro, K. A. (2024). Il faut repenser la place de la compétence numérique dans le système éducatif québécois. *Médiations Et médiatisations*, (17), 108–114. https://doi.org/10.52358/mm.vi17.385

Gouvernement du Québec. (2023). *Priorités du ministre Drainville : 81 M$ pour la formation professionnelle*. https://www.quebec.ca/nouvelles/actualites/details/priorites-du-ministre-drainville-81-m-pour-la-formation-professionnelle-45734

Gouvernement du Québec. (2019). *Cadre de référence de la compétence numérique*. https://www.education.gouv.qc.ca/fileadmin/site_web/documents/ministere/Cadre-reference-competence-num.pdf

Haidt, J. (2023). Get Phones Out of Schools Now[29]. *The Atlantic* https://www.theatlantic.com/technology/archive/2024/03/teen-childhood-smartphone-use-mental-health-effects/677722/?gift=6kW-3kpyH7tV4NcHqyTUk4pQbLwhJ51mzwtSdzMDnc4&utm_source=copy-link&utm_medium=social&utm_campaign=share

Institut national en santé publique du Québec (INSPQ). (2024). Écrans et hyperconnectivité. https://www.inspq.qc.ca/ecrans

Jokhan, A., Sharma, B., et Singh, S. (2019). Early warning system as a predictor for student performance in higher education blended courses. *Studies in Higher Education*, 44(11), 1900-1911. https://doi.org/10.1080/03075079.2018.1466872

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., et Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial* Intelligence, 3, 100074.

Langlais, J.F., Mosconi, E. et Guillemette, M. G. (2023). Developing a Data Literacy Assessment Tool for the Business Workforce. *AMCIS 2023 Proceedings. 7.* https://aisel.aisnet.org/amcis2023/fow/fow/7

La Presse canadienne. (2023, 27 décembre). L'interdiction du cellulaire en classe en vigueur après les Fêtes au Québec. https://ici.radio-canada.ca/nouvelle/2037612/interdiction-cellulaire-ecole-classes-quebec

Le Crosnier, H., Ertzscheid, O., Peugeot, V., Mercier, S., Berthaud, C., Charnay, D. et Maurel, L. (2011). Vers les « communs de la connaissance ». Documentaliste-Sciences de l'Information, 48, 48-59. https://doi.org/10.3917/docsi.483.0048

Massou, L. et Lavielle-Gutnik, N. (2017). Enseigner à l'université avec le numérique: Savoirs, ressources, médiations. De Boeck Supérieur. https://doi.org/10.3917/dbu.masso.2017.01

Massou, L. (2022). Mutualisation des ressources pédagogiques numériques pour l'hybridation : vers l'éducation ouverte ?. Distances et médiations des savoirs. OpenEdition. https://doi.org/10.4000/dms.7997

Ministère de l'Éducation du Québec. (2023) Plan stratégique 2023-2027. https://cdn-contenu.quebec.ca/cdn-contenu/adm/min/education/publications-adm/plan-strategique/plan-strategique-MEQ-2023-2027.pdf

Ministère de l'Éducation du Québec. (2024). Programme d'études *Culture et citoyenneté québécoise*. https://www.education.gouv.qc.ca/parents-et-tuteurs/references/refonte-programme-ethique-culture-religieuse/

Ministère des Affaires étrangères de Finlande. (s.d). Pour des décisions éclairées : l'éducation finlandaise aux médias prévient la désinformation. *finland.fi.* https://finland.fi/fr/vie-amp-societe/pour-des-decisions-eclairees-leducation-finlandaise-aux-medias-previent-la-desinformation/

Ministry of Education and Culture of Finland. (s.d.). Media literacy and education. https://okm.fi/en/media-literacy

Mosconi, E., Hidous, M. et Jorge, L.F. (2024). Enhancing Digital Literacy for Managers and Leaders through a Self-Assessment Tool. AMCIS 2024 Proceedings.

Naffi, N. et Beaulieu, N. (2023, décembre). *Responsables de la formation en santé : Réinventer l'apprentissage pour mieux soigner – Rapport d'analyse des besoins.* CHUM-Obvia. https://www.obvia.ca/ressources/responsables-de-la-formation-en-sante-reinventer-lapprentissage-pour-mieux-soigner

Naffi, N., Charest, M., Danis, S., Pique, L., Davidson, A., Brault, N., Bernard, M. et Barma, S. (2023). Empowering Youth to Combat Malicious Deepfakes and Disinformation : An Experiential and Reflective Learning Experience Informed by Personal Construct Theory. *Journal Of Constructivist Psychology*, 1-22. https://doi.org/10.1080/10720537.2023.2294314

Ontario. (2024). L'Ontario restreint l'utilisation des cellulaires et interdit le vapotage dans les écoles. https://news.ontario.ca/fr/release/1004501/lontario-restreint-lutilisation-des-cellulaires-et-interdit-le-vapotage-dans-les-ecoles.

Pacheco, A., Da Silva, C. G. et De Freitas, M. C. V. (2023). A metadata model for authenticity in digital archival descriptions. *Archival Science, 23*(4), 629-673.

Papadoudi-Ros, H. (2020). Luc Massou, Brigitte Juanals, Philippe Bonfils, Philippe Dumas (dirs), *Sources ouvertes numériques. Usages éducatifs, enjeux communicationnels*: Nancy, PUN - Éditions universitaires de Lorraine, 2019, 286 pages. Questions de communication, 38, 671-675. https://doi.org/10.4000/questionsdecommunication.24715

Planchon, R. (2023). «En Suède, les tout-petits sont exposés dès la crèche aux écrans». *Le Figaro*. https://www.lefigaro.fr/vox/societe/en-suede-les-tout-petits-sont-exposes-des-la-creche-aux-ecrans-20230526

Psyché, V. et Ruer, P. (2019). L'apprentissage adaptatif intelligent. *Revue Le Tableau, 8*(4).

Plantard, P., Clerget, J., Bléhaut, M. et Mansencal, L. B. (2023). Lever le voile sur l'éloignement du numériques: approches conceptuelles et quantification du phénomène en France.

Poellhuber, B., Marquis, C., Roy, N. et M. Linck (2024). ChatGPT au postsecondaire: perceptions et usages. Webinaire présenté à l'invitation de l'AQPC, le 26 avril 2024.

29 Version texte gratuit du même article: https://www.afterbabel.com/p/phone-free-schools?r=182klo&utm_campaign=post&utm_medium=email

*Projet de loi C-63: Loi édictant la Loi sur les préjudices en ligne, modifiant le Code criminel, la Loi canadienne sur les droits de la personne et la Loi concernant la déclaration obligatoire de la pornographie juvénile sur Internet par les personnes qui fournissent des services Internet et apportant des modifications corrélatives et connexes à d'autres lois.* (2024). 1re sess., 42e lég. https://www.parl.ca/LegisInfo/fr/projet-de-loi/44-1/c-63

Richard, C. (2018). Dans la boîte noire des algorithmes: Comment nous nous sommes rendus calculables. Revue du Crieur, 11, 68-85. https://doi.org/10.3917/crieu.011.0068

Sen, A. K., (1999). Development as Freedom. New York, Anchor Books, 2000.

Sghir, N., Adadi, A. et Lahmer, M. (2023). Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and information technologies*, 28(7), 8299-8333.

World Economic Forum. (2024). *Global Risks Report 2024.*https://www.weforum.org/publications/global-risks-report-2024/digest/

Zheng, B., Warschauer, M., Lin, C.-H. et Chang, C. (2016). Learning in One-to-One Laptop Environments: A Meta-Analysis and Research Synthesis. *Review of Educational Research*, *86*(4), 1052-1084. https://doi.org/10.3102/0034654316628645

## Axe Industrie 4.0, travail et emploi

Autor, D. H. (2019). Work of the Past, Work of the Future. In *AEA Papers and Proceedings* (Vol. 109, pp. 1-32). https://doi.org/10.1257/pandp.20191110

Autor, D. H. et Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *The American Economic Review*, *103*(5), 1553-1597.

Berg, J., Green, F., Nurski, L. et Spencer, D. A. (2023). Risks to job quality from digital technologies: Are industrial relations in Europe ready for the challenge? *European Journal of Industrial Relations*, *0*(0), 1-19.

Campa, R. (2017). Technological unemployment : a brief history of an idea. *ISA eSymposium for Sociology*, *7*(1), 1-16. http://www.sagepub.net/isa/admin/viewEBPDF.aspx?&art=EBul-Campa-Mar2017.pdf

Carbonell, J. S. (2022). *Le futur du travail.* Éditions Amsterdam.

Conseil de l'innovation du Québec. (2024). Vers un Québec innovant : Rapport de recommandations pour repenser le modèle québécois d'appui à la recherche et développement des entreprise, Avril 2024

de Marcellis-Warin N. (2022). Analyse comparative d'écosystèmes en IA dans le but de repérer les pratiques innovantes en matière de formation et de transfert de connaissances. (2022RP-20). https://doi.org/10.54932/SXOH392

Dostie B. et Dufour G. (2023). *Stratégie de mise à niveau technologique des entreprises québécoises.* CIRANO. https://cirano.qc.ca/fr/sommaires/2023RP-01

Dupuis, M, G. Murray et M. Wey (2022). Les transformations numériques et l'ordre négocié des techniciens des services automobiles au Québec. *Négociations*, 2 (38), 73-97.

Doellgast, V., et Wagner, I. (2022). Collective regulation and the future of work in the digital economy: Insights from comparative employment relations. *Journal of Industrial Relations*, 64(3), 438-460. https://doi.org/10.1177/00221856221101165

Ford, M. (2015). *Rise of the Robots. Technology and the threat of a jobless future.* Basic Books.

Frey, C. B., et Osborne, M. (2013). *The Future of Employment* (O. M. P. o. T. a. Employment, Ed.) [Working Paper]. University of Oxford

Frey, C. B., et Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254-280.

Friedmann, G. (1956). *Le travail en miettes. Spécialisation et loisirs.* Éditions de l'Université de Bruxelles.

Garneau, J., Pérez-Lauzon, S. et Lévesque, C. (2023). Digitalisation of work in aerospace manufacturing: expanding union frames and repertoires of action in Belgium, Canada and Denmark. *Transfer: European Review of Labour and Research, 49*(1)

Gesualdi-Fecteau, D., Murray, G., Lévesque, C. et Roby, N. (2023). What makes work better?. *Transfer: European Review of Labour Research, (forthcoming).*

Gmyrek, P., Berg, J., et Bescond, D. (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality. *ILO Working Paper*, *96*.

Jacob, S., Souissi, S et Trudel, J.S. (2021). *Intelligence artificielle et transformation du métier de gestionnaire.* Chaire de recherche sur l'administration publique à l'ère numérique (Université Laval). https://www.administration-numerique.chaire.ulaval.ca/sites/administration-numerique.chaire.ulaval.ca/files/uploads/IA%20et%20m%C3%A9tier%20gestionnaire.pdf

Kellogg, K. C, Valentine, M. A, et Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366-410.

Lévesque, C., Bowkett, C., Garneau, J. M. É., et Perez-Lauzon, S. (2024). Crafting human-centred AI in workspaces for better work. In J.-L. Denis, C. Régis, A. Kishimoto, et M. Axente (Eds.), *Human-Centered AI: Multidisciplinary Perspectives for Policy-Makers, Auditors and Users.* Taylor & Francis Group, LLC.

Naville, P. (1961). *L'automation et le travail humain. Rapport d'enquête (France 1957-1959).* Centre National de la Recherche Scientifique.

## Axe Arts, médias et diversité culturelle

Académie de la transformation numérique et Centre d'études sur les médias. (2023) *Vers la maturité numérique des entreprises de presse.* Rapports de recherche remis au ministère de la Culture et des Communications du Québec. Université Laval. https://www.cem.ulaval.ca/publications/vers-la-maturite-numerique-des-entreprises-de-presse-repertoire-de-pratiques-numeriques-innovantes/ [30]

Agence France-Presse. (2024, 16 janvier). Sora, le nouveau générateur de vidéos d'OpenAI, inquiète médias et artistes, *Radio Canada*, en ligne : https://ici.radio-canada.ca/nouvelle/2050048/sora-generateur-ia-openai

Arguedas, A. R., et Simon, F. M. (2023). *Automating democracy: Generative AI, journalism, and the future of democracy.* Balliol College, University of Oxford.

Arielli, E. et Monovich, L. (2021-2024) *Artificial Aesthetics: Generative AI, art and visual media*n, en ligne : http://manovich.net/index.php/projects/artificial-aesthetics

Azzaria G. (2023). Droit d'auteur et intelligence artificielle : un aperçu canadien dans Castets-Renard, C. et Eynard. J (dir.), *Un droit de l'intelligence artificielle. Entre règles sectorielles et régime général. Perspectives comparées.* Bruylant.

Baillargeon S. (2022, 31 décembre). L'IA sème le doute dans le milieu de l'art, *Le Devoir*, en ligne : https://www.ledevoir.com/culture/776236/intelligence-artificielle-l-ia-seme-le-doute-dans-le-milieu-de-l-art

Béasse, M. (2021). Intelligence artificielle et mesures d'audience du journalisme en ligne: pour quelle intelligence des publics? *Les Cahiers du journalisme-Recherches* 2(7), R25-R38.

Beaudoin L., Duhaime C., Guèvremont V. et Taillon, P. (2024). *La souveraineté culturelle du Québec à l'ère du numérique : Rapport du comité-conseil sur la découvrabilité des contenus culturels.* https://cdn-contenu.quebec.ca/cdn-contenu/adm/min/culture-communications/publications-adm/rapport/RA-comite-expert-decouvrabilite-contenus-culturels.pdf

Beckett, C. et Yaseen, M. (2023). *Générer le changement : Enquête mondiale sur l'utilisation de l'IA par les organismes de presse.* London School of Economics. https://static1.squarespace.com/static/64d60527c01ae7106f2646e9/t/6582087ffa08ac19291268ec/1703020692847/Generating-Change-_-The-Journalism-AI-report-_-French.pdf

Brin, C. et Guèvremont, V. (dir.). (2024). *Intelligence artificielle, culture et médias.* Presses de l'Université Laval

Cardon, D. (2019). *Culture numérique.* Paris. Presses de Sciences Po.

Brin, C., Charlton, S. et Marois, A. (2023). *Digital News Report Canada. Synthèse des données 2023,* juin.Centre d'études sur les médias (CEM).

https://www.cem.ulaval.ca/publications/dnr-2023-canada-fr/

Tremblay, F. et Charlton, S. (2023). *Regard sur les pratiques d'information au Canada.* Digital News Report 2023, juin. https://www.cem.ulaval.ca/publications/dnr-canada-2023-regard/

Fallaha, S., Langlois, L. et Maclure, J. (2024). Réflexion collective sur l'encadrement de l'IA. Dossier thématique no 5. *Les autres impacts sociétaux de l'intelligence artificielle.* Conseil de l'innovation du Québec.

https://conseilinnovation.quebec/intelligence-artificielle/publications-de-la-reflexion-collective/

Crimaldi, F. et Leonelli, M. (2023). AI and the creative realm: A short review of current and future applications, en ligne : https://www.researchgate.net/publication/371310981_AI_and_the_creative_realm_A_short_review_of_current_and_future_applications

Culture Montérégie. (2023). Sujet de l'heure, l'IA et les changements qu'elle amène en art, en ligne : https://culturemonteregie.qc.ca/lintelligence-artificielle-et-la-culture/#:~:text=L'IA%20a%20%C3%A9galement%20eu%20un%20impact%20sur%20la%20pr%C3%A9servation,peintures%20et%20des%20artefacts%20historiques

Dandurand, G., Blottière, M., Jorandon, G., Gertler, N., Wester, M., Chartier-Edward, N., Roberge, J. et McKelvey, F. (2023). *Entraîner l'actualité: La couverture canadienne du cycle d'engouement pour l'IA (2012-2021).* INRS - Urbanisation Culture Société. https://espace.inrs.ca/id/eprint/13227/

De Cremer, D., Bianzino, N. M. et Falk, B. (2023) How generative AI could disrupt creative work, *Harvard Business Review*, en ligne : https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work

Dierickx, L. (2024). La régulation reste humaine. *Revue Projet*, (1), 40-43. https://cdjm.org/journalisme-et-intelligence-artificielle-les-bonnes-pratiques/

Eapen, T. T., Finkenstadt, D. J., Folk, J., et Venkataswamy, L. (2023). How Generative AI Can Augment Human Creativity: Use it to promote divergent thinking, *Harvard Business Review*, 101:4, 56-64, en ligne : https://hbr.org/2023/07/how-generative-ai-can-augment-human-creativity

30 Ce lien mène vers un des quatre documents de cette série, mais les trois autres documents sont toutefois accessibles à partir de cette page.

Faljaoui, A. (2024). Des journaux de référence comme *Le Monde* ou le *Financial Times* ont-ils vendu leur âme en donnant accès à leurs articles à OpenAI, la maison mère de ChatGPT ? *RTBF Actus,* 14 mai. https://www.rtbf.be/article/des-journaux-de-reference-comme-le-monde-ou-le-financial-times-ont-ils-vendu-leur-ame-en-donnant-acces-a-leurs-articles-a-openai-la-maison-mere-de-chatgpt-11370159

Forum économique mondial. (2024). *Global Risks Report 2024.*. https://www.weforum.org/publications/global-risks-report-2024/digest/

Giroux, D. (2022). *Les médias québécois d'information. État des lieux en 2022.* Centre d'études sur les médias. https://www.cem.ulaval.ca/publications/etat-des-lieux-en-2022/

Hildén, J. 2022. The public service approach to recommender systems: Filtering to cultivate. *TV & New Media* 23:7, 777-796.

Jung, B. (2023). *The Dynamics of the Machine Heuristic, Misinformation, and Political Informedness* (Doctoral dissertation, University of Kansas), en ligne: https://www.proquest.com/docview/2863283639?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

Koltermann, F. (2020). Le photojournalisme à l'ère de la post-vérité. *European Journalism Observatory*. https://fr.ejo.ch/deontologie-qualite/avenir-photojournalisme-post-verite

Labrecque, A. (2023). Invasion de faux sites de nouvelles générés par l'intelligence artificielle, *Québec Science*, en ligne. https://www.quebecscience.qc.ca/technologie/faux-sites-nouvelles-intelligence-artificielle/

Metz, C. (2024, 16 janvier), Des vidéos à couper le souffle générées par l'IA, *La Presse*, en ligne : https://www.lapresse.ca/affaires/techno/2024-02-16/nouveau-logiciel-d-openai/des-videos-a-couper-le-souffle-generees-par-l-ia.php

Mons, B. Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B. et Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information services & use*, *37*(1), 49-56. https://content.iospress.com/articles/information-services-and-use/isu824

Parker, S., Park, S., Pehlivan, Z., Abrahams, A., Desblancs, M., Owen, T., Phillips, J. et Bridgman, A. (2024). *When journalism is turned off: Preliminary findings on the effects of Meta's news ban in Canada.* Media Ecosystem Observatory, McGill University. https://www.mediaecosystemobservatory.com/reports/state-of-the-canadian-media-ecosystem-ldra7

Pettineo, C. (2024). Les médias face à l'intelligence artificielle: 20 chartes passées au crible. *La Revue des médias.* Institut national de l'audiovisuel (INA), 11 janvier. https://larevuedesmedias.ina.fr/les-medias-face-lintelligence-artificielle-20-chartes-passees-au-crible

Radio France. (2022, 4 février). Juice est une nouvelle application qui permet de créer et d'écouter des playlists simplement avec la voix, en ligne : https://www.radiofrance.fr/francebleu/podcasts/connecte/juice-est-une-nouvelle-application-qui-permet-de-creer-et-d-ecouter-des-playlists-simplement-avec-la-voix-3846643

RFA Tibetan. (2023, 15 novembre). New Tibetan translation software harnesses AI to preserve language, *Radio Free Asia*, en ligne : https://www.rfa.org/english/news/tibet/translation-software-11152023122728.html

Rioux, C. (2024, 16 janvier), Le milieu culturel se mobilise pour protéger le droit d'auteur à l'ère de l'IA, Radio Canada, en ligne : https://ici.radio-canada.ca/nouvelle/2042119/droit-auteur-intelligence-artificielle-consultation-publique

Runco, M. A. (2023). AI can only produce artificial creativity. *Journal of Creativity*, 33(3) https://doi.org/10.1016/j.yjoc.2023.100063

Siegel, S. (2024). The Phenomenal Public. *Political Philosophy* 1(1). https://philpapers.org/rec/SIETPP

Simon, F. M. et Isaza-Ibarra, L. F. (2023). *AI in the news: reshaping the information ecosystem ?* Oxford Internet Institute, University of Oxford.

https://www.oii.ox.ac.uk/news-events/reports/ai-in-the-news-reshaping-the-information-ecosystem/

Stray, J. (2021). Making artificial intelligence work for investigative journalism. *Algorithms, Automation, and News*, 97-118.

Wingström, R., Hautala, J., et Lundman, R. (2024). Redefining Creativity in the Era of AI? Perspectives of Computer Scientists and New Media Artists, *Creativity Research Journal*, 36(2). https://doi.org/10.1080/10400419.2022.2107850

## Axe Sobriété numérique et transition socio-écologique

Abrassart C. (2011). *La naissance de l'éco-conception : acteurs, raisonnements, enjeux de pilotage et horizons d'une rationalisation industrielle (1990-2010).* Thèse soutenue à l'École Nationale Supérieure des Mines de Paris.

Abrassart C., Roche S., Begue F et Gourret A. (2024). Une cartographie des enjeux et des problématisations de la sobriété numérique. Colloque ACFAS 29 - Numérique responsable : état des lieux et perspectives, Ottawa, 15 mai 2024.

ADEME. (2022), *En route vers la sobriété numérique*. https://librairie.ademe.fr/ged/6555/guide-en-route-vers-sobriete-numerique-202209.pdf

ADEME. (s.d). Calculateur de l'empreinte carbone des usages numériques. https://agirpourlatransition.ademe.fr/particuliers/bureau/numerique/calculez-lempreinte-carbone-usages-numeriques

ADEME. (2024). Transitions 2050. Lien : https://www.ademe.fr/les-futurs-en-transition/

ADEME et Arcep. (2023). *Étude sur l'empreinte environnementale du numérique en 2020, 2030 et 2050.*https://www.arcep.fr/la-regulation/grands-dossiers-thematiques-transverses/lempreinte-environnementale-du-numerique/etude-ademe-arcep-empreinte-environnemental-numerique-2020-2030-2050.html

Arcep, Arcom et ADEME. (2024). *Référentiel général de l'écoconception des services numériques.* https://www.arcep.fr/actualites/actualites-et-communiques/detail/n/environnement-rgesn-170524.html

Aggeri F., Beulque R. et Micheaux H. (2023). *L'économie circulaire.* Repères, La Découverte, Paris.

Bigo A. et Heran, F. (2020). Malus poids, émissions de $CO_2$ : intéressons-nous enfin aux véhicules intermédiaires !. *The Conversation.* https://theconversation.com/malus-poids-emissions-de-co-interessons-nous-enfin-aux-vehicules-intermediaires-148650#:~:text=En%20septembre%202020%2C%20le%20minist%C3%A8re,d'1%2C8%20tonne.

Bonamy, C., Boudinet, C., Bourgès, L., Dassas, K., Lefèvre, L., Ninassi, B., & Vivat, F. (2022). L'écoconception d'un service numérique: des actions pour réduire l'impact environnemental du numérique. *1024: Bulletin de la Société Informatique de France*, (19), 59-68. https://hal.science/hal-03660118/document

Cézard, F., et Mourad, M. (2019). *Panorama sur la notion de sobriété. Définitions, mises en œuvre, enjeux.* ADEME. Panorama sur la notion de sobriété - La librairie ADEME

Crawford K. (2021). *Atlas of AI,* Yale University Press.

Designers Éthiques. (2022). *Guide éco-conception des services numériques* https://eco-conception.designersethiques.org/guide/fr/

Durand Folco J. et Martineau J. (2023), *Le capital algorithmique*, Éditions *écosociété*, Montréal.

Eco-Index. (s.d). Quel est l'impact environnemental de votre site ? https://www.ecoindex.fr

Flipo, F. (2017). Peut-on croire aux TIC « vertes » ?. *Annales des Mines - Responsabilité et environnement*, 87, 105-107.

Flipo, F. (2021). L'impératif de la sobriété numérique. *Cahiers Droit, Sciences & Technologies*, (13), 29-47.

Hache, E. (2022), « La sobriété, vecteur de puissance? », Numéro thématique Géopolitique de la sobriété, *Revue internationale et stratégique*, (4), 77-86.

Heikkilä M. (2023). Making an image with generative AI uses as much energy as charging your phone. MIT Technology Review, December 1, 2023.

Hoang, A. N., Mellot, S. et Prodhomme (2022). Le numérique à l'épreuve de l'écologie: La fabrique de la « sobriété numérique » comme nouvelle norme? *Interfaces numériques*, *11*(1).

ISO. (2020). ISO 14006:2020-Systèmes de management environnemental — Lignes directrices pour intégrer l'éco-conception. https://www.iso.org/fr/standard/72644.html

Lascoumes P. et Le Galès P. (2005). *Gouverner par les instruments*, Presses de Sciences Po, Paris.

LINC. (2023). *Cahier IP9 – Données, empreinte environnementale et libertés.* Lien : https://linc.cnil.fr/cahier-ip9-donnees-empreinte-et-libertes

IPCC. (2022). *Climate Change 2022. Mitigation of Climate Change. Full Report.* https://www.ipcc.ch/report/sixth-assessment-report-working-group-3/

Lopez F. (2022). *À bout de flux.* Éditions divergences, Paris.

Luccioni A.S., Jernite Y. et Strubell, E. (2023). Power Hungry Processing: Watts Driving the Cost of AI Deployment? *arXiv preprint.* https://arxiv.org/abs/2311.16863

Ministère de l'Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs (2023). *Stratégie gouvernementale de développement durable 2023-2028.* Québec, 130 pages.

Monnin A. et Ben Kemoun M. (2022). La sobriété comme suffisance intensive. L'exemple de la musique. CNM Lab – Dossier *La musique en mouvements. Horizon 2030.* https://cnmlab.fr/recueil/horizon-la-musique-en-2030/chapitre/8/

Nova N. et Roussilhe G. (2020). Du low-tech numérique aux numériques situés. *Sciences du Design*, (1), 91-101. https://www.cairn.info/revue-sciences-du-design-2020-1-page-91.htm

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., So, D., Texier, M. et Dean, J. (2021). Carbon emissions and large neural network training. *arXiv preprint.* https://arxiv.org/pdf/2104.10350

Riggs, W., et Shukla, S. (2021). Exploring the benefits of minimobility in the urban context. *Journal of Transport and Land Use*, *14*(1), 1019-1037.

RTE. (2002), *Futurs énergétiques 2050.* Rapport complet. https://www.rte-france.com/analyses-tendances-et-prospectives/bilan-previsionnel-2050-futurs-energetiques

Shift Project. (2020), *Déployer la sobriété numérique.* https://theshiftproject.org/article/deployer-la-sobriete-numerique-rapport-shift/

Shift Project. (2021), *Impact environnemental du numérique : tendances à 5 ans et gouvernance de la 5G.*https://theshiftproject.org/article/impact-environnemental-du-numerique-5g-nouvelle-etude-du-shift/

UNEP (2023), *Emissions Gap Report 2023*. https://www.unep.org/resources/emissions-gap-report-2023

UNITAR (2024), *The Global E-waste Monitor 2024*.https://ewastemonitor.info/the-global-e-waste-monitor-2024/

Villalba B. (2022). *Politiques de sobriété*, Le Pommier, Paris.

## Axe Droit, cyberjustice et cybersécurité

Albert-Rochette, M., Lavis, C. et Meunier, M. (2023). Protection de la vie privée dans le contexte des plateformes en ligne : les limites des fiducies de données et du devoir fiduciaire de loyauté. *Lex Electronica, 28*(1), 27-59. https://lexelectronica.openum.ca/files/sites/103/Protection-de-la-vie-privee-dans-le-contexte-des-plateformes-en-ligne_Albert-Rochette_Lavis-et-Meunier-4-7.pdf

Autorité des marchés financiers. (2024). *Document de réflexion et de discussion – Meilleures pratiques pour l'utilisation responsable de l'IA dans le secteur financier*. https://lautorite.qc.ca/fileadmin/lautorite/grand_public/publications/professionnels/tous-les-pros/IA_DocumentReflexion_MeilleuresPratiques2024.pdf

Balkin, J. M. (2020). The Fiduciary Model of Privacy. *Harvard Law Review Forum, 134*, 11-33. https://harvardlawreview.org/wp-content/uploads/2020/10/134-Harv.-L.-Rev.-F.-11.pdf

Beaudoin, L., Duhaime, C., Guèvremont, V. et Taillon, P. (2024). *La souveraineté culturelle du Québec à l'ère du numérique : Rapport du Comité-conseil sur la découvrabilité des contenus culturels*. Comité-conseil sur la découvrabilité des contenus culturels (ministère de la Culture et des Communications du Québec). https://cdn-contenu.quebec.ca/cdn-contenu/adm/min/culture-communications/publications-adm/rapport/RA-comite-expert-decouvrabilite-contenus-culturels.pdf

Benthall, S. et Shekman, D. (2023). Designing Fiduciary Artificial Intelligence. *Proceedings of ACM Conference (Conference'17)*. https://arxiv.org/pdf/2308.02435

Boine, C. (2023, 29 septembre). *Fiduciary Law to Promote Value Alignment in AI Systems* [présentation d'une conférencière invitée, We Robot 2023]. Faculté de droit, Boston University. https://www.bu.edu/law/files/2023/09/Fiduciary-paper.pdf

Botes, M. (2023). Regulating scientific and technological uncertainty: The precautionary principle in the context of human genomics and AI. *South African Journal of Science, 119*(5/6). https://doi.org/10.17159/sajs.2023/15037

Centre québécois du droit de l'environnement. (2024). *Principe de précaution et principe de prévention : quelle est la différence ?* https://cqde.org/fr/sinformer-nouvelle/les-bases-du-droit-de-lenvironnement/principe-precaution-principe-prevention-difference/

*Charte des droits et libertés de la personne*. RLRQ, c. C-12. https://www.legisquebec.gouv.qc.ca/fr/document/lc/C-12

*Code civil du Québec*. RLRQ, c. CCQ-1991. https://www.legisquebec.gouv.qc.ca/fr/document/lc/ccq-1991

Conseil de l'Innovation du Québec. (2024). *Prêt pour l'IA – Répondre au défi du développement et du déploiement responsables de l'IA au Québec*. https://conseilinnovation.quebec/wp-content/uploads/2024/02/Rapport_IA_CIQ-1.pdf

Department for Science, Innovation and Technology of the United Kingdom. (2023, 3 août). *Policy paper – A pro-innovation approach to AI regulation*. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#executive-summary

EU Artificial Intelligence Act. (2024). *Les textes de loi – Projet final (2024)*. https://artificialintelligenceact.eu/fr/l-acte/

Gautrais, V., Tchiniaev, A. et Guiraud, É. (2023a, février). *Guide des bonnes pratiques en intelligence artificielle : Sept principes pour une utilisation responsable des données*. Chaire L.R. Wilson sur le droit des technologies de l'information et du commerce électronique, Centre de recherche en droit public et Observatoire international sur les impacts sociétaux de l'IA et du numérique (Obvia). https://www.obvia.ca/sites/obvia.ca/files/ressources/202307-OBV-Pub-Guide_BP_UtilisationDonn%C3%A9es.pdf

Gautrais, V., Trudel, P. et Vermeys, N. (2023b, 28 novembre). *LCCJTI+ : perspectives de mise à jour de la Loi concernant le cadre juridique des technologies de l'information (RLRQ c C-1.1) 2001-2023 – Rapport final – 28 novembre 2023*. Chaire L.R. Wilson sur le droit des technologies de l'information et du commerce électronique et Centre de recherche en droit public. https://www.crdp.umontreal.ca/files/sites/101/2024/02/Rapport-final.pdf

Hulin, A.-S. (2021). De la fiducie de données en droit civil québécois : étude exploratoire pour un outil en construction. *McGill Law Journal, 67*(2), 119-156. https://doi.org/10.7202/1093269ar

Lacroix, M. et Vermeys, N. (2023). *RESPONSABILITÉ.IA – Du droit québécois de la responsabilité civile à l'égard de l'intelligence artificielle*. Éditions Yvon Blais.

Larivière-Mantha, S. et Hulin, A.-S. (2024). *Réflexion collective sur l'encadrement de l'IA – Dossier thématique no1 – Le cadre de gouvernance de l'intelligence artificielle*. Conseil de l'innovation du Québec. https://conseilinnovation.quebec/wp-content/uploads/2024/02/T1-DossierThematique-3.pdf

*Loi concernant le cadre juridique des technologies de l'information*. RLRQ, c. C-1.1. https://www.legisquebec.gouv.qc.ca/fr/document/lc/c-1.1

*Loi modifiant la Loi sur la gouvernance et la gestion des ressources informationnelles des organismes publics et des entreprises du gouvernement et d'autres dispositions législatives*. LQ. (2023). C. 28. https://www.publicationsduquebec.gouv.qc.ca/fileadmin/Fichiers_client/lois_et_reglements/LoisAnnuelles/fr/2023/2023C28F.PDF

*Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels*. RLRQ, c. A-2.1. https://www.legisquebec.gouv.qc.ca/fr/document/lc/a-2.1

*Loi sur la gouvernance et la gestion des ressources informationnelles des organismes publics et des entreprises du gouvernement*. RLRQ, c. G-1.03. https://www.legisquebec.gouv.qc.ca/fr/document/lc/G-1.03

*Loi sur la protection des renseignements personnels dans le secteur privé*. RLRQ, c. P-39.1. https://www.legisquebec.gouv.qc.ca/fr/document/lc/p-39.1#:~:text=Les%20renseignements%20personnels%20concernant%20un,25%2C%20a

*Loi sur les renseignements de santé et de services sociaux*. RLRQ, c. R-22.1. https://www.legisquebec.gouv.qc.ca/fr/document/lc/R-22.1

Marchildon, A., Boine, C., Sabourin Laflamme, A., Anctil, D., Boudreau LeBlanc, A., Auclair, S., Balagué, C., Morin-Bertrand, F.-A., Castets-Renard, C., Jackson, P. et Langlois, L. (2024). *Un an après l'arrivée de ChatGPT : Réflexions de l'Obvia sur les enjeux et pistes d'action possibles face à l'IA générative*. Obvia. https://doi.org/10.61737/ZKWZ3721

Parlement européen. (2024, 16 avril). *Rectificatif du Parlement européen du 16 avril 2024*. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_FR.pdf

*Projet de loi C-27 : Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois*. (2022). 1re sess., 44e lég. https://www.parl.ca/legisinfo/fr/projet-de-loi/44-1/c-27?view=progress

*Règlement sur l'anonymisation des renseignements personnels*, D. 783-2024, (2024) 156 G.O.Q. II, 2847

Richards, N. M. et Hartzog, W. (2021). A Duty of Loyalty for Privacy Law. *Washington University Law Review, 99*, 961-1021. https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=4055&context=faculty_scholarship

### Bibliographie indicative complémentaire:

Bernatchez, S. (2023). L'action normative. La dynamique du droit en mouvement. Dans M. Hautereau-Boutonnet et C. Sintez (dir.), *Être en recherche. Études en l'honneur du professeur Catherine Thibierge* (p. 23-38). Éditions Mare &Martin.

Bernatchez, S., Bouchard, A. et Bélanger, S.-M. (2021). Le droit de la gouvernance pour réguler la gouvernance algorithmique. *Éthique Publique, 23*(2). https://doi.org/10.4000/ethiquepublique.6559

Castets-Renard, C. et Eynard, J. (dir.). (2023). *Un droit de l'intelligence artificielle – Entre règles sectorielles et régime général, perspectives comparées*. Bruylant.

Conseil de l'innovation du Québec. (2023, octobre). *Réflexion collective sur l'encadrement de l'IA – Le cadre de gouvernance de l'intelligence artificielle : portrait du cadre normatif applicable au Québec – État de situation*. https://conseilinnovation.quebec/wp-content/uploads/2023/10/CIQ_cadre_gouvernance_IA_EDS-1.pdf

Gautrais, V. (2012). *Neutralité technologique : rédaction et interprétation des lois face aux changements technologiques*. Éditions Thémis.

Latil, A. (2023). *Le droit du numérique – Une approche par les risques*. Dalloz.

Lehaire, B. (2022). *L'innovation hors-la-loi – Les origines de la techno-normativité* (1ère éd.). Larcier.

## Axe Éthique, gouvernance et démocratie

Amugongo, L. M., Kriebitz, A., Boch, A. et Lütge, C. (2023). Operationalising AI ethics through the agile software development lifecycle: a case study of AI-enabled mobile health applications. *AI and Ethics*, 1-18.https://link.springer.com/article/10.1007/s43681-023-00331-3

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management, 62*. https://doi.org/10.1016/j.ijinfomgt.2021.102433

Auclair, S. (2023). Penser la protection des renseignements inférés sur la santé des personnes : du contrôle individuel à l'encadrement des utilisations en fonction de leur légitimité. *Éthique Publique, 25*(1). https://doi.org/10.4000/ethiquepublique.7979

Bélisle-Pipon, J.-C., Monteferrante, E., Roy, M.-C., et Couture, V. (2022b). Artificial intelligence ethics has a black box problem. *AI & SOCIETY : Journal of Knowledge, Culture and Communication, 38*(4), 1507–1522. https://doi.org/10.1007/s00146-021-01380-0

Bélisle-Pipon, J.-C., Ravitsky, V., Bensoussan, Y., et Bridge2AI-Voice Consortium. (2023). Individuals and (Synthetic) Data Points: Using Value-Sensitive Design to Foster Ethical Deliberations on Epistemic Transitions. *The American Journal of Bioethics : AJOB, 23*(9), 69–72. https://doi.org/10.1080/15265161.2023.2237436

Bélisle-Pipon, J.-C., Couture, V., et Roy, M.-C. (2022a). Equality and Equity in Compensating Patient Engagement in Research: A Plea for Exceptionalism. *Research Ethics, 18*(2), 126–131. https://doi.org/10.1177/17470161211059993

Bengio, Y. (2023). AI and Catastrophic Risk. *Journal of Democracy, 34*(4), 111–121. https://doi.org/10.1353/jod.2023.a907692

Bernatchez, S., Bouchard, A., & Bélanger, S.-M. (2021). Le droit de la gouvernance pour réguler la gouvernance algorithmique. *Éthique Publique, 23*(2). https://doi.org/10.4000/ethiquepublique.6559

Besse, P. (2020). Détecter, évaluer les risques des impacts discriminatoires des algorithmes d'IA. https://hal.science/hal-02616963/

Besse P. (2021). Anticiper les Risques Juridiques des Systemes d'IA. https://www.math.univ-toulouse.fr/~besse/pub/AnticipeRisquesJuridIA.pdf

Bietti, E. (2020, Janvier). From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy. Dans *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 210-219). https://doi.org/10.1145/3351095.3372860

Boch, A., Lucaj, L. et Corrigan, C. (2021). *A Robotic New Hope: Opportunities, Challenges, and Ethical Considerations of Social Robots.* Institute for Ethics in Artificial Intelligence (IEAI). https://ieai.sot.tum.de/wp-content/uploads/2021/05/ResearchBrief_April2021_SocialRobots_FinalV2.pdf

Boddington, P. (2023). *AI ethics : a textbook*. Springer. https://doi.org/10.1007/978-981-19-9382-4

Boudreau LeBlanc, A. et Williams-Jones, B. (2023). Chapitre 6. Des éthiques collectives à une gestion adaptative des conflits organisationnels L'outil co P·R·I·M·O·V en gouvernance. *Journal International de Bioéthique et d'Éthique Des Sciences, 34*(3), 103. https://doi.org/10.3917/jibes.343.0103

CACM Staff. (2012). Operationalizing privacy by design. *Communications of the ACM, 55*(9), 7. https://doi.org/10.1145/2330667.2330669

Conseil du statut de la femme. (2023). Intelligence artificielle : des risques pour l'égalité entre les femmes et les hommes. https://csf.gouv.qc.ca/wp-content/uploads/Avis_intelligence_artificielle.pdf

Dauvergne, P. (2022). Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. *Review of International Political Economy, 29*(3). https://doi.org/10.1080/09692290.2020.1814381

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines : Journal for Artificial Intelligence, Philosophy and Cognitive Science, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Gentelet, K. et Bahary-Dionne, A. (2021). Stratégies des Premiers Peuples au Canada concernant les données numériques: décolonisation et souveraineté. *tic&société, 15*(1| 1er semestre 2021), 189-208.. https://doi.org/10.4000/ticetsociete.6063

Gentelet, K. (dir.) (2023). *Les intelligences artificielles au prisme de la justice sociale. Considering Artificial Intelligence Through the Lens of Social Justice*. Les Presses de l'Université Laval. https://www.pulaval.com/livres/les-intelligences-artificielles-au-prisme-de-la-justice-sociale-considering-artificial-intelligence-through-the-lens-of-social-justice

Gentelet, K. et Mizrahi, S. (2024). A Human-Centered Approach to AI Governance: Operationalizing Human Rights Through Civil Participation. Dans Régis, C., Denis, J.L., Axente, M.L. et Kishimoto, A., *Human-Centered AI: a Multidisciplinary Perspective for Policy-Makers, Auditors and Users. CRC Press, Florida, sous-presse* https://www.taylorfrancis.com/chapters/edit/10.1201/9781003320791-24/human-centered-approach-ai-governance-karine-gentelet-sarit-mizrahi

Gentelet, K., Mathieu, L.C. et Bahary-Dionne, A. (2024). *Le numérique et l'intelligence artificielle comme outil de justice sociale : les initiatives par et pour les groupes et communautés marginalisé.es*. Obvia. https://doi.org/10.61737/JICE5848

Helbing, D., Fanitabasi, F., Giannotti, F., Hänggli, R., Hausladen, C. I., van den Hoven, J., ... Mahajan, S., Pedreschi, D. et Pournaras, E. (2021). Ethics of smart cities: Towards value-sensitive design and co-evolving city life. *Sustainability, 13*(20), 11162. https://doi.org/10.3390/su132011162

Hulin, A.-S. (2022). La gouvernance des données en droit civil québécois : comment (re)concilier protection et exploitation des données personnelles ? *Revue Internationale de Droit Économique, XXXVI*(3), 39. https://doi.org/10.3917/ride.363.0039

Hulin, A.-S., Guiraud, É., Lawarée, J., et Langlois, L. (2023). Le partage et la mise en commun des données de santé : quels enjeux pour un objectif d'innovation sociale responsable ? *Éthique Publique, 25*(1). https://doi.org/10.4000/ethiquepublique.7983

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, 1(9), 389-399. https://www.nature.com/articles/s42256-019-0088-2

Jacob, S. et Lawarée, J. (2022). *Les mesures publiques dans les stratégies gouvernementales en matière d'IA : une perspective internationale*. Obvia. https://doi.org/10.61737/HNRV6126

Khawaja, Z., et Bélisle-Pipon, J.-C. (2023). Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. *Frontiers in Digital Health, 5*, 1278186. https://doi.org/10.3389/fdgth.2023.1278186

Koniakou, V. (2023). From the "rush to ethics" to the "race for governance" in Artificial Intelligence. *Information Systems Frontiers*, 25(1), 71-102. https://doi.org/10.1007/s10796-022-10300-6

Langlois, L., & Castets-Renard, C. (2021). Présentation - La gouvernance algorithmique. *Éthique Publique, 23*(2). https://doi.org/10.4000/ethiquepublique.6574

Larivière-Mantha, S. et Hulin, A.-S. (2024). *Réflexion collective sur l'encadrement de l'IA – Dossier thématique no1 – Le cadre de gouvernance de l'intelligence artificielle*. Conseil de l'innovation du Québec. https://conseilinnovation.quebec/wp-content/uploads/2024/02/T1-DossierThematique-3.pdf

Lee, J. A., Hilty, R. et Liu, K. C. (dir.). (2021). *Artificial intelligence and intellectual property*. Oxford University Press. OU

Li, J., Almentero, B. K., et Besse, C. (2022). Explainable machine learning for drug shortage prediction in a pandemic setting. Dans *International Conference on Machine Learning, Optimization, and Data Science* (pp. 141-155). Cham: Springer Nature Switzerland. https://link.springer.com/chapter/10.1007/978-3-031-25599-1_11

McLennan, S., Fiske, A., Tigard, D., Müller, R., Haddadin, S., & Buyx, A. (2022). Embedded ethics: a proposal for integrating ethics into the development of medical AI. *BMC Medical Ethics*, *23*(1), 6. https://doi.org/10.1186/s12910-022-00746-3

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature machine intelligence*, *1*(11), 501-507. https://doi.org/10.1038/s42256-019-0114-4

Marchildon, A., Jacob, S. et Marcoux, A.M. (2021). *Grille de réflexivité sur les enjeux éthiques des systèmes d'intelligence artificielle (SIA)*. Obvia. https://doi.org/10.61737/OXHA5372

Marchildon, A., Martin, D. et Martineau, J.T. (2024a). Démarche de conception du Cadre d'éthique en intelligence artificielle de l'administration publique québécoise. *Ministère de la Cybersécurité et du Numérique*. https://www.quebec.ca/gouvernement/politiques-orientations/vitrine-numeriqc/comprendre-et-suivre/evenements/evenement-public-numeriqc/demarche-de-conception-du-cadre-dethique-en-intelligence-artificielle-de-ladministration-publique-quebecoise; https://www.youtube.com/watch?v=_IXTJ_02p2Q&t=2085s

Marchildon, A., Boine, C., Sabourin Laflamme, A., Anctil, D., Boudreau LeBlanc, A., Auclair, S., Balagué, C., Morin-Bertrand, F.-A., Castets-Renard, C., Jackson, P. et Langlois, L. (2024b). *Un an après l'arrivée de ChatGPT : Réflexions de l'Obvia sur les enjeux et pistes d'action possibles face à l'IA générative*. Obvia. https://doi.org/10.61737/ZKWZ3721

Marty, F. (2019). La protection des algorithmes par le secret des affaires. *Revue internationale de droit économique*, (2), 211-237. https://www.cairn.info/revue-internationale-de-droit-economique-2019-2-page-211.htm?WT.tsrc=cairnEmailAlert&WT.mc_id=RIDE_332

Meyers, J.M. (2018). Artificial intelligence and trade secrets. *Landslide*, *11*, 17. https://heinonline.org/HOL/LandingPage?handle=hein.journals/lndslid11&div=40&id=&page=

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Osmanlliu, E., Paquette, J., Grenier, A. D., Lewis, P., Bouthillier, M. E., Bédard, S. et Pomey, M. P. (2022). Fantastic perspectives and where to find them: involving patients and citizens in digital health research. *Research Involvement and Engagement*, *8*(1), 37. https://doi.org/10.1186/s40900-022-00374-6

Petitgand, C.(2024). *Données personnelles : reprenons le pouvoir! : réflexions sur la gouvernance citoyenne à l'ère numérique*. Presses de l'Université Laval. https://www.pulaval.com/livres/donnees-personnelles-reprenons-le-pouvoir#:~:text=Le%20pouvoir%20des%20donn%C3%A9es%20n,du%20pouvoir%20sur%20nos%20donn%C3%A9es.

Poussart, D. (2021). Gouvernance algorithmique, intelligence artificielle, enjeux, éthique: esquisse d'une analyse critique. «D'après vous en tant que chercheur en IA, quels sont les enjeux d'une gouvernance algorithmique? L'éthique a-t-elle un rôle à jouer?». *Éthique publique. Revue internationale d'éthique sociétale et gouvernementale*, *23*(2) https://doi.org/10.4000/ethiquepublique.6418

Seger, E. (2022). In defence of principlism in AI ethics and governance. *Philosophy & Technology*, 35(2), 45. https://doi.org/10.1007/s13347-022-00538-y

Shapiro S.S. (2010). Privacy by design: Moving from art to practice. *Communications of the ACM*, *53*(6), 27–29. https://doi.org/10.1145/1743546.1743559

Spelda, P. et Stritecky, V. (2020). The future of human-artificial intelligence nexus and its environmental costs. Futures, 117, 102531.

Tidjon, L. N. et Khomh, F. (2022). The Different Faces of AI Ethics Across the World: A Principle-To-Practice Gap Analysis. *IEEE Transactions on Artificial Intelligence*, 1-20. https://doi.org/10.1109/TAI.2022.3225132

UNESCO. (2023). UNESCO's Recommendation on the Ethics of Artificial Intelligence: key facts. https://unesdoc.unesco.org/ark:/48223/pf0000385082

Ursin, F., Lindner, F., Ropinski, T., Salloch, S., et Timmermann, C. (2023). Levels of explicability for medical artificial intelligence: What do we normatively need and what can we technically reach?. *Ethik in der Medizin*, *35*(2), 173-199. https://link.springer.com/article/10.1007/s00481-023-00761-x

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. Dans *Proceedings of the 2019 AAAI/ACM Conference*

Zakir, M. H., Khan, S. H., & Saeed, Z. (2023). The Impact of Artificial Intelligence on Intellectual Property Rights. *International journal of human and society*, *3*(4), 312-319. https://ijhs.com.pk/index.php/IJHS/article/view/330

## Bibliographie complémentaire indicative (publications des membres de l'axe)


Anctil, D. (2023). L'éducation supérieure à l'ère de l'IA générative. *Pédagogie collégiale vol. 36, no 3, printemps-été 2023*.

Dubé, S., Santaguida, M., Anctil, D., Leshner, C., Vachon, D., et Johnson, A. (2021). Technology-based sexualities. *Encyclopedia of sexuality and gender*, 1-12.

Dubé, S., Santaguida, M., Anctil, D., Zhu, C. Y., Thomasse, L., Giaccari, L., ... Johnson, A. (2023). Perceived stigma and erotic technology: From sex toys to erobots. *Psychology & Sexuality*, *14*(1), 141–157. https://doi.org/10.1080/19419899.2022.2067783

Déziel, Pierre-Luc and Auclair, Sylvain and Sirois, Caroline and Khoury, Richard and Simard, Marc, Applying Contextual Integrity to the Development of Polypharmacy Indicators in Canada (July 1, 2023). 5th Annual Symposium on Applications of Contextual Integrity (2023), Available at SSRN: https://ssrn.com/abstract=4555409 or http://dx.doi.org/10.2139/ssrn.4555409

Bélanger, D., et Silvey, R. (2020). An Im/mobility turn: power geometries of care and migration. *Journal of Ethnic and Migration Studies*, *46*(16), 3423-3440.

Bouhouita-Guermech, S., Gogognon, P., et Bélisle-Pipon, J. C. (2023). Specific challenges posed by artificial intelligence in research ethics. Frontiers in artificial intelligence, 6.

Shevlane, T., Farquhar, S., Garfinkel, B., Phuong, M., Whittlestone, J., Leung, J., ... et Dafoe, A. (2023). Model evaluation for extreme risks. *arXiv preprint arXiv:2305.15324*.

Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Harari, Y. N., ... et Mindermann, S. (2023). Managing ai risks in an era of rapid progress. *arXiv preprint arXiv:2310.17688*.

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., ... et Anderljung, M. (2020). Toward trustworthy AI development: mechanisms for supporting verifiable claims. *arXiv preprint arXiv:2004.07213*.

Bergeron, M. (2022). Évaluation éthique d'applications de traçage : perspective Québécoise. *Droit, Santé et Société*, 2-3, 82-86.https://www.cairn.info/revue-droit-sante-et-societe-2022-2-page-82.htm.

Bernatchez, S. (2020). De la démocratie par le droit à la dictature des algorithmes? La théorie juridique à l'ère cybernétique. *Lex Electronica*, *25*(3), 10-33

Bernatchez, S. (2023). Une révolution scientifique en droit? Les fondements cybernétiques des théories juridiques contemporaines. *Revue interdisciplinaire d'études juridiques*, *90*(1), 3-41.

Bernatchez, S. (2023). Une révolution scientifique en droit ? Les fondements cybernétiques des théories juridiques contemporaines. *Revue interdisciplinaire d'études juridiques*, 90, 3-41. https://doi.org/10.3917/riej.090.0003

Besse, P., del Barrio, E., Gordaliza, P., Loubes, J. M., et Risser, L. (2022). A survey of bias in machine learning through the prism of statistical parity. *The American Statistician*, *76*(2), 188-198.

Besse, P., Besse-Patin, A., et Castets-Renard, C. (2020). Implications juridiques et éthiques des algorithmes d'intelligence artificielle dans le domaine de la santé. *Statistique et Société*, *8*(3), 21-53

Castets-Renard, C., et Besse, P. (2022). Ex Ante accountability of the AI Act: Between certification and standardization, in pursuit of fundamental rights in the country of compliance. *Pursuit of Fundamental Rights in the Country of Compliance (August 29, 2022). Artificial Intelligence Law: Between Sectoral Rules and Comprehensive Regime. Comparative Law Perspectives, C. Castets-Renard et J. Eynard (eds), Bruylant, Forthcoming*.

Besse, P. (2022). Conformité européenne des systèmes d'IA: outils statistiques élémentaires. *Statistique et société*, (10| 3), 25-46.

Besse, P. Statistique & Risques Juridiques (à venir) des Systèmes d'IA

Hohma, E., Boch, A., et Trauth, R. (2022). Towards an Accountability Framework for Artificial Intelligence Systems

Boch, A., et Corrigan, C. (2020). Ethics and the use of AI-based tracing tools to manage the COVID-19 pandemic. *TUM IEAI Research Brief*, *2020*

Lütge, C., Hohma, E., Boch, A., Poszler, F., et Corrigan, C. (2022). On a Risk-Based Assessment Approach to AI Ethics Governance

Hulin, A. S. (2021). De la fiducie de données en droit civil québécois: étude exploratoire pour un outil en construction. McGill Law Journal, 67(2), 119-156.

obvia

obvia.ca