

Les impacts sociétaux de l'intelligence artificielle sur la démocratie, l'environnement et les arts et la culture

État de situation

Octobre 2023

Avec la participation financière de

Québec 



Table des matières

Remerciements	2
Impacts sociétaux de l'intelligence artificielle.....	3
I. Démocratie et intelligence artificielle	3
1. IA, processus électoral et communications entre le gouvernement et les citoyens	4
2. IA, délibération publique et sphère médiatique	8
3. Encadrement du rôle de l'IA dans le processus démocratique	13
Bibliographie	15
II. Impact environnemental et intelligence artificielle.....	21
1. Conséquences environnementales de l'IA	21
2. Rôle des normes internationales	28
3. Survol d'initiatives d'encadrement hors Québec	29
Annexe	30
Bibliographie	31
III. Intelligence artificielle, arts et culture	34
1. Diffusion et consommation des contenus culturels	34
2. Création des contenus culturels	36
3. Perception et compréhension des contenus culturels.....	39
4. Survol des initiatives internationales.....	40
Bibliographie	42

Remerciements

Cet état de situation en trois parties a été produit grâce à l'appui de la Commission de l'éthique en science et en technologie (CEST), l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA) et le Centre d'expertise international de Montréal en intelligence artificielle (CEIMIA).

Nous souhaitons remercier tout particulièrement François Boucher, David Hughes, Janick Houde et Félix-Arnaud Morin-Bertrand.



Impacts sociaux de l'intelligence artificielle

L'intelligence artificielle (IA) est une technologie d'usage général dont la pénétration dans toutes les sphères de l'activité humaine, incluant au sein de la société québécoise, générera des effets aussi bien positifs que négatifs. Le Québec a commencé à en étudier les impacts, à élaborer ou mobiliser les connaissances qui serviront à les prendre en compte, et à utiliser des outils que les organisations pourront appliquer pour développer ou déployer l'IA de manière responsable. Toutefois, beaucoup de travail reste à faire pour mieux comprendre – et mieux faire comprendre – l'ensemble des enjeux soulevés par la montée de l'IA et les stratégies à élaborer qui permettront de les relever efficacement.

Dans ce qui suit, nous présentons un état de situation visant à cerner les principaux défis découlant de l'impact de l'IA sur la démocratie, l'environnement ainsi que la culture et les arts au Québec. Pour chacun de ces thèmes, nous cherchons à identifier les principaux risques et avantages de l'IA tout en proposant un survol des tendances et initiatives existantes en matière d'encadrement de l'IA.

I. Démocratie et intelligence artificielle

La démocratie repose sur l'idée que le peuple est l'auteur des lois et qu'il se gouverne collectivement. Les démocraties modernes ont mis en place différents systèmes permettant à des représentants élus d'incarner la voix du peuple et de contrôler le processus législatif. La démocratie, telle qu'on la connaît maintenant, présuppose ainsi, minimalement, la tenue d'élections et la possibilité pour tout citoyen de faire campagne pour se faire élire. Cependant, la démocratie ne saurait se réduire à la simple procédure formelle électorale. Elle s'incarne également dans des processus informels, plus diffus, de délibération publique, par lesquels l'opinion et la volonté des citoyens se forment et se transmettent aux représentants et aux institutions du gouvernement.

Nous avons choisi de ne pas nous limiter à une vision minimaliste de la démocratie dans laquelle celle-ci ne reposerait que sur la tenue d'élections à intervalles réguliers permettant de remplacer les représentants élus qui n'arrivent pas à traduire les préférences de la majorité en matière de politiques publiques. Une telle conception de la démocratie est associée à la conception dite « agrégative¹ », autrefois défendue par l'économiste Joseph Schumpeter.

Sans remettre en question la centralité du processus électoral, les travaux plus récents sur la démocratie soulignent sur l'importance de la délibération publique en tant que mécanisme essentiel assurant la légitimité des décisions et institutions politiques², promouvant la qualité de la prise de décisions politiques³ et permettant aux citoyens de contester ces dernières⁴. Cette conception élargie de la démocratie structure notre analyse : nous cherchons à dresser un portrait global des risques et des promesses liés à l'IA vis-à-vis des différentes composantes de la démocratie représentative et délibérative.

Dans un premier temps, nous nous pencherons sur les promesses et les menaces liées à l'usage de l'IA dans le processus électoral. Dans un second temps, nous nous intéresserons aux impacts de l'IA sur la délibération citoyenne, laquelle s'opère à l'intérieur d'une sphère médiatique, soit un espace de communication entre la société civile et l'État, sphère qui est maintenant largement numérique et structurée par les réseaux sociaux et Internet. Enfin, nous proposerons un survol de l'encadrement des systèmes d'intelligence artificielle (SIA) qui ont une incidence sur le processus démocratique dans des sociétés modernes similaires au Canada.

¹Schumpeter (2008).

²Habermas (1997) et Rawls (1995).

³Estlund (2009) et Hélène Landemore (2017).

⁴Pettit (2000) et Tully (2008).



Bien que nous adoptions une conception élargie de la démocratie, nous devons néanmoins limiter notre démarche afin de respecter les échéanciers associés au mandat confié au Conseil de l'innovation du Québec (CIQ), de sorte que les deux thèmes ci-dessous liés à la démocratie en sont exclus :

- L'usage de l'IA pour bonifier l'accès aux services publics et rendre ceux-ci plus efficaces (thème retenu dans certaines synthèses sur l'IA et la démocratie [Thiel, 2022] ou monographies [Duberry, 2022b]);
- L'incidence de l'IA sur les « valeurs fondamentales des démocraties » au sens large (incluant par exemple les inégalités socioéconomiques, la discrimination, l'autonomie personnelle)⁵.

1. IA, processus électoral et communications entre le gouvernement et les citoyens

En tant que forme de gouvernement par le peuple, la démocratie presuppose l'existence d'institutions et de procédures permettant d'une part aux citoyens de communiquer leurs préférences sociales et politiques et, d'autre part, au gouvernement de consulter ces derniers afin de connaître leurs préférences. Dans les démocraties représentatives modernes, le processus électoral, qui inclut au moins le vote et la possibilité de faire campagne pour persuader les électeurs, joue un rôle central dans ce processus de communication et d'agrégation des préférences des citoyens. Une littérature émergente s'est constituée autour des promesses de l'IA pour améliorer ce processus électoral.

1.1 Avantages de l'IA dans le processus électoral

Certains soulignent que des outils alimentés par l'IA pourraient augmenter le taux de participation aux élections. Depuis les élections présidentielles américaines de 2016, on a largement eu recours à des robots conversationnels, tels que HelloVote, pour solliciter, avec un certain succès, la participation électorale des citoyens par le biais d'envois automatisés de rappels et d'information relative aux modalités du vote (où, quand, comment, etc.)⁶. On peut raisonnablement s'attendre à ce que l'automatisation des efforts pour solliciter le vote avec l'usage de robots conversationnels alimentés par de grands modèles de langage (*Large Language Models*) augmente l'efficacité de ces campagnes⁷.

L'usage de tels robots par les partis politiques et les candidats aux élections pourrait également permettre à ceux-ci de communiquer plus efficacement avec l'électorat, pour mieux faire connaître leur offre politique et les valeurs qu'ils défendent. On pourrait même envisager la possibilité pour les électeurs d'interagir avec le jumeau numérique des candidats⁸.

Une telle technique permettrait de personnaliser le lien entre candidats et électeurs et de créer un sentiment de proximité. Les jumeaux numériques sont des représentations digitales d'objets réels. Ils sont construits à partir de la collecte de vastes ensembles de données relatives à ces objets et d'analyses des régularités et tendances observées

⁵ Nous nous pencherons néanmoins sur l'impact qu'a l'IA sur l'égalité politique des citoyens (c.-à-d. la capacité à prendre part de manière égale aux processus de décisions collectives), sur la manière dont la discrimination algorithmique affecte cette participation et sur l'autonomie politique des citoyens entendue comme capacité à exprimer leurs propres choix et opinions politiques. Pour une synthèse de ces enjeux, voir Maclure et Saint-Pierre (2018).

⁶ Scola (2016).

⁷ Sanders (2023).

⁸ *Ibid.*



dans ces données, permettant ainsi de prédire le comportement de l'objet modélisé dans différentes situations et de le reproduire de manière réaliste⁹.

La personnalisation des relations entre l'électorat, d'un côté, et les gouvernements, les élus, les partis ou les candidats de l'autre, peut également s'appliquer à l'électorat. En effet, les techniques de microciblage actuelles permettent de segmenter la population en groupes de personnes partageant certains traits les rendant plus susceptibles de réagir d'une manière donnée à certains messages. Le microciblage est « une technique de marketing politique visant la création de messages conçus spécifiquement pour un individu, au moyen d'analyses produites à partir des données personnelles de cette personne en fonction de critères tels que les caractéristiques démographiques, le style de vie, les habitudes de consommations et bien d'autres »¹⁰.

Aux États-Unis, ces techniques sont omniprésentes dans les campagnes présidentielles depuis 2004. Au Canada, elles le sont depuis 2006 dans les campagnes fédérales¹¹. Elles permettraient notamment aux partis de mieux interroger les électeurs sur les sujets qui les intéressent vraiment et de diversifier leur discours pour rendre l'offre des partis plus inclusive et plus représentative de la population¹².

Les gouvernements et les partis politiques peuvent également ajuster leur offre politique en sondant la population non pas en établissant des profils à partir de données numériques (ce qui relève du microciblage), mais en déployant des techniques d'écoute sociale qui consistent à analyser les débats publics sur les réseaux sociaux¹³. Les algorithmes alimentés par l'apprentissage automatique ont la capacité de traiter en continu des ensembles massifs de données constitués de millions de publications sur les réseaux sociaux. Cela permet d'établir certaines classifications révélant les thèmes et tendances qui reflètent les préoccupations des citoyens. Certains vont même jusqu'à suggérer que le double numérique des citoyens se prononce de façon automatisée sur toutes les questions susceptibles d'affecter les gens dans le monde réel. Ce modèle est parfois nommé « démocratie des avatars¹⁴ » ou « démocratie augmentée¹⁵ ».

On envisage aussi ce qu'on pourrait appeler l'« écoute sociale structurée », qui est un exemple frappant des avantages qu'offre l'IA lorsqu'il s'agit de mieux consulter la population. L'idée est d'observer et d'analyser les interventions des participants sur des plateformes numériques structurées de manière à produire certains résultats¹⁶.

Par exemple, le gouvernement de Taiwan a mis en place une plateforme numérique, vTaiwan, alimentée par le système Pol.is¹⁷. Cette plateforme permet aux citoyens d'exprimer leur opinion sur des questions de politiques publiques, de publier une proposition sur un sujet donné (p. ex. : « il faut interdire les hypertrucages » [deepfakes en anglais]) et d'attribuer une mention positive ou négative aux propositions des autres (« je suis d'accord » ou « je ne suis pas d'accord »). L'avantage de la plateforme est que les citoyens ne peuvent pas commenter les propositions des autres, ce qui évite les interactions toxiques de type « trolling ». De ces opinions se dégagent ensuite des orientations consensuelles¹⁸; des algorithmes classent les propositions similaires et permettent de visualiser les données sur des cartes virtuelles afin de situer les usagers. Ceux-ci peuvent ensuite reformuler leurs propositions de sorte à obtenir

⁹ Pour une revue de la littérature sur les jumeaux numériques, voir Jones et autres (2020).

¹⁰ Borgesius et autres (2018), Caron (2023), Kreiss (2017) et Lavigne (2022).

¹¹ Caron (2023).

¹² Borgesius et autres (2018). Pour des discussions critiques sur l'efficacité du microciblage politique, voir aussi Gorton (2016) et Tappin et autres (2023).

¹³ Duberry (2022a).

¹⁴ Himmelreich (2022).

¹⁵ Hidalgo (2019).

¹⁶ Pour en savoir plus sur l'impact que peut avoir la manière dont la structure des plateformes numériques se base sur les comportements, attitudes et croyances des citoyens, voir Simons (2023).

¹⁷ <https://pol.is/home>.

¹⁸ Miller (2019).



plus de mentions positives. Or, on a observé qu'en s'adaptant à ces cartes dynamiques et en reformulant leurs propositions, les usagers finissent par converger vers des propositions plus nuancées et plus précises¹⁹.

1.2 Risques démocratiques de l'IA dans le processus électoral

Les techniques de rédaction automatisée, de microciblage et d'écoute sociale à grande échelle comportent néanmoins d'importants risques pour la démocratie. Comme nous le verrons dans la prochaine section, certaines de ces applications peuvent avoir un effet nocif sur la qualité de la délibération publique. Pour l'instant, nous nous contenterons de souligner comment ces techniques peuvent miner les procédures et institutions qui permettent la communication entre gouvernements et citoyens.

Tout d'abord, les gains en matière de participation électorale et de sondages des préférences qui découlent de ces techniques sont inégalement répartis. En effet, ces gains ont peu de chance de bénéficier aux personnes ayant un faible niveau d'éducation ou de littératie numérique, ou encore à celles qui sont laissées pour compte à cause de la fracture numérique (*digital divide*) et de la fracture des données (*data divide*)²⁰. Les gens qui ont du mal à exprimer clairement leur pensée à l'écrit ou qui n'utilisent pas les réseaux sociaux risquent en effet de ne pas recevoir les messages des partis politiques et, ainsi, d'être invisibles pour les systèmes d'écoute sociale de masse.

Par ailleurs, il n'est pas du tout certain que les personnes connectées le soient dans toutes les sphères de leur vie et qu'elles soient les seules à utiliser leurs objets connectés et comptes personnels. Ainsi, une grande lectrice qui achète ses livres dans une librairie de quartier en payant en espèces ne verra pas cette dimension de son existence apparaître dans la constitution de son double numérique. Pour cette raison, rien n'indique clairement que les techniques de profilage et de classification permettant le microciblage et l'écoute sociale de masse arrivent à construire un portrait fidèle des usagers²¹.

La démocratie ne se limite pas à la prise en compte des préférences des citoyens. Elle exige également que chacun ait une capacité égale d'influencer les décisions politiques²². Si cette égalité ne doit pas être seulement formelle, le processus politique ne doit pas non plus être influencé par les inégalités sociales et économiques²³. Or, les techniques d'écoute sociale et de microciblage politique sont susceptibles de renforcer certaines inégalités.

Comme on l'a souligné précédemment, la voix des personnes peu connectées et produisant peu de données ressort moins dans les résultats des algorithmes d'apprentissage automatique. Il en va de même pour la voix des personnes qui appartiennent à différents groupes sociaux minoritaires, et ce, même lorsqu'elles produisent des données. Les algorithmes d'apprentissage automatique ont en effet tendance à reproduire les discriminations existantes à l'égard des personnes appartenant à des minorités sociales. Cela se produit lorsqu'ils appliquent des conclusions statistiques valides pour des groupes ou des sous-catégories de personnes à des cas individuels en ignorant certaines spécificités pertinentes, lorsqu'ils se nourrissent de données reflétant elles-mêmes des injustices, notamment des données reflétant la surcriminalisation de certains groupes sociaux²⁴, ou encore lorsque ces minorités sont moins présentes dans les données d'entraînement²⁵.

¹⁹ Horton (2018). Pour une analyse détaillée des risques et opportunités démocratiques de l'usage de Pol.is avec de grands modèles de langage, voir C. T. Small et autres (2023).

²⁰ La fracture numérique réfère aux inégalités d'accès à Internet et aux technologies numériques, tandis que la fracture des données en est une sous-catégorie. Elle désigne les inégalités liées au fait que les personnes qui produisent moins de données (p. ex. : parce qu'elles utilisent moins d'objets connectés, voire pas du tout) sont moins bien servies ou représentées par les algorithmes alimentés par l'apprentissage automatique.

²¹ Gorton (2016).

²² Christiano (2010).

²³ Rawls (2008) et Rosanvallon (2011).

²⁴ Mattu (2016).

²⁵ Maclure et Saint-Pierre (2018); Mittelstadt et autres (2016).



Une critique plus radicale de l'impact de l'IA sur l'égalité politique souligne que le modèle de marketing politique fondé sur l'analyse de données massives s'inspire de la dynamique du « capitalisme de la surveillance²⁶ » ou du « capitalisme algorithmique²⁷ », et qu'il la renforce. Dans cette nouvelle phase de l'économie capitaliste, les entreprises, principalement les géants du numérique, compétitionnent pour obtenir l'attention des usagers de leurs plateformes et pour extraire le plus de valeur des capacités de prédiction et d'influence libérées par l'analyse automatisée des données massives. Selon cette critique, le recours au marketing politique numérique alimenté par l'IA favorise ainsi la concentration de pouvoir dans les mains d'acteurs privés (et souvent étrangers) qui ont des intérêts tout autres que la santé de la démocratie et le bien-être du public québécois. Les récentes difficultés du gouvernement fédéral à réguler les plateformes numériques en raison du refus d'obtempérer de géants américains nous rappellent que la capacité d'autodétermination des États peut être entravée par des acteurs privés puissants, étrangers ou non.

Bien que la critique du capitalisme de surveillance algorithmique tende à cibler l'influence d'acteurs privés, le recours par des acteurs publics à ce modèle de gouvernance soulève autant d'enjeux démocratiques. D'une part, on peut se demander si ce modèle compromet l'égalité politique en avantageant les grands partis déjà établis qui ont accès à plus de ressources pour collecter et analyser des données massives ou pour recourir à des services disposant de cette expertise. Il faut néanmoins soupeser ce risque à la lumière de la démocratisation de l'usage des grands modèles de langage qu'on observe actuellement. Cette démocratisation pourrait jouer en faveur des partis plus petits ou émergents en « niveler le terrain de jeu ».

D'autre part, l'usage par l'État de techniques d'écoute sociale, de profilage et de prédiction soulève des questions en lien avec le droit à la vie privée qui sont pertinentes pour la qualité de la vie démocratique. Par exemple, ces techniques pourraient permettre à des États de cibler des mouvements et des discours de contestation légitime. D'une manière plus générale, on pourrait même dire que le simple fait de se sentir observé modifie le comportement d'une manière qui peut mettre en péril l'autonomie, personnelle aussi bien que politique, de même que la liberté d'expression et de conscience²⁸. En outre, la protection de la vie privée est intimement liée à l'égalité politique en ceci que le secret du vote est nécessaire pour protéger les électeurs de certaines formes d'ostracisme sociales et politiques²⁹. Les capacités prédictives et de profilage actuelles de l'IA pourraient même permettre de prédire d'une manière assez précise le vote des personnes de telle sorte que cela affaiblirait la protection contre de telles formes d'ostracisme³⁰.

Enfin, pour certains, l'usage à grande échelle du marketing politique numérique, notamment du microciblage, engendre potentiellement un déficit de transparence. En divisant finement les électeurs en segments exposés à différents messages adaptés à leurs profils, les partis sont susceptibles de rendre invisibles différents aspects de leur programme et de leur discours. De plus, de manière générale, plusieurs travaux sur l'IA et la démocratie se sont concentrés sur le déficit de légitimité lié au problème de l'explicabilité et de l'opacité des algorithmes³¹. En effet, comme les algorithmes d'apprentissage automatique sont opaques, même aux yeux de leurs propres développeurs, une ombre d'ilégitimité plane sur toutes les décisions assistées par de tels systèmes puisque la légitimité démocratique exige minimalement que les décisions publiques contraignantes soient soutenues par des raisons (justifications) accessibles à tous les citoyens. Plus particulièrement, si les plateformes électorales des partis politiques sont largement conçues sur la base de recommandations issues d'algorithmes prédictifs, on pourrait soutenir que ni les électeurs ni les candidats n'ont accès aux raisons qui fondent les orientations politiques qui sont proposées par les différents partis³².

²⁶ Shoshana Zuboff (2020).

²⁷ Folco et Martineau (2021).

²⁸ Reiman (1995).

²⁹ Lever (2012).

³⁰ Mainz et autres (2022).

³¹ Beckman et autres (2022); Chomanski (2022); Danaher (2016); de Fine Licht (2020); Maclure (2021).

³² Fung et Lessig (2023).



2. IA, délibération publique et sphère médiatique

Les institutions et pratiques sociales permettant aux citoyens d'exprimer leurs préférences politiques par l'entremise d'élections visant à élire des représentants qui exercent une fonction législative sont au cœur des systèmes démocratiques modernes. Toutefois, les théories contemporaines de la démocratie attirent également notre regard en amont de cette pratique, sur la manière dont les préférences politiques se forment par la délibération publique. On entend ici toutes les conversations et tous les échanges de raison à propos des affaires publiques qui ont lieu entre les citoyens, dans les médias et à l'intérieur des groupes composant la société civile (associations, universités, syndicats, organisations culturelles ou religieuses, etc.).

Dans cette vision plus riche de la démocratie, la délibération publique est valorisée : en tant qu'élément nécessaire à la légitimité du pouvoir politique; en tant que procédure qui éclaire et informe mieux les décisions collectives; en tant qu'espace nécessaire à la contestation du pouvoir politique. La délibération peut également être un mode d'apprentissage pour les citoyens qui s'y engagent puisqu'ils sont ainsi exposés aux idées des autres et peuvent modifier leurs positions selon la réfutation de certaines hypothèses³³. Dans les pages suivantes, nous identifierons d'abord les principales menaces posées par l'IA aux trois aspects de la délibération : légitimité et justification publique; valeur épistémique de la délibération; contestation du pouvoir politique. Puis, nous nous pencherons sur les manières dont l'IA peut bonifier la délibération publique.

Avant d'aborder l'incidence de l'IA sur la délibération, soulignons le déficit délibératif qui pourrait résulter de la mise en place d'une certaine vision « techno-enthousiaste » de la gouvernance démocratique algorithmique³⁴. Il importe de prendre au sérieux, comme nous l'avons souligné, les avantages potentiels de l'IA quant à la consultation citoyenne et à la transmission des préférences politiques individuelles au gouvernement. Toutefois, il serait problématique que prévalse une conception de la démocratie entièrement fondée sur l'écoute sociale massive et automatisée. Une telle conception s'appuierait alors sur une vision passive des citoyens comme étant de simples émetteurs de données, tout en abandonnant l'aspect délibératif de la démocratie et en se privant des gains ayant trait à la légitimité et à la rationalité de la contestation.

2.1 Légitimité, justification publique et opacité

Plusieurs travaux théoriques contemporains sur la démocratie soulignent le caractère essentiel de la délibération publique pour la justification des politiques publiques et la légitimité de l'exercice du pouvoir politique. Ils insistent sur l'importance de clarifier les paramètres de la délibération et de déterminer un ensemble de valeurs et de principes publics suscitant l'adhésion de la grande majorité des membres de la société. Dans les sociétés modernes traversées par des désaccords profonds sur les questions ultimes liées au sens de la vie, à la religion, à la justice et à l'éthique, il importe de baser les décisions publiques sur des valeurs politiques communes, plutôt que sur des doctrines controversées générant des désaccords fondamentaux³⁵.

2.2 Valeur épistémique de la délibération

D'autres travaux sur la démocratie délibérative accordent moins d'importance à la nécessité de s'appuyer sur un inventaire public de raisons communes pour justifier les décisions politiques. Ces travaux se concentrent plutôt sur la qualité du processus de délibération ainsi que sur les conditions qui font obstacle à la bonne délibération ou qui permettent à la délibération de mener à de meilleures décisions collectives. Une vision optimiste de la délibération tend à vanter ses avantages épistémiques, c'est-à-dire ses bénéfices sur les plans de la connaissance, de la qualité de la cognition et de la rationalité. L'idée est que la libre discussion expose à des points de vue diversifiés et permet de corriger des erreurs factuelles et des préjugés, tout en rendant disponibles les connaissances et les savoirs

³³ Habermas (1987).

³⁴ Ouellet (2021).

³⁵ Gutmann et Thompson (2004); Rawls (2005).



expérientiels des autres membres de la société³⁶. Les décisions collectives prises à l'issue de telles délibérations seraient ainsi plus rationnelles et plus éclairées³⁷.

À l'heure actuelle, la délibération publique se fait en grande partie dans la sphère médiatique. Cette dernière est largement constituée de plateformes numériques et de médias sociaux qui sont alimentés par des algorithmes visant à recommander des contenus aux usagers, mais aussi par des contenus produits par des IA génératives (robots conversationnels, rédaction humaine assistée par de grands modèles de langage, contenus audio et vidéo, etc.)³⁸. Les travaux récents sur l'impact de l'IA sur la délibération publique tendent à sonner l'alarme sur les risques que l'IA exacerbe certains dysfonctionnements de la délibération publique, par exemple en favorisant la polarisation et la manipulation par le biais de la désinformation.

2.2.1 Polarisation

Des critiques de la démocratie délibérative ont déjà remis en question la vision optimiste selon laquelle la délibération mène nécessairement à des résultats plus rationnels et éclairés. Il arrive par exemple que des mécanismes économiques, psychologiques ou sociaux génèrent des obstacles à la bonne délibération³⁹. Le phénomène de la polarisation politique des groupes est un exemple d'échec de la délibération. Accentué par les réseaux sociaux alimentés par l'IA, ce phénomène se produit lorsque les opinions prédélibératives des membres se renforcent suite à la délibération au sein de ces groupes⁴⁰. Les phénomènes de chambre d'écho et de bulle épistémique (ou « bulle de filtre ») contribuent à la polarisation. Une chambre d'écho émerge lorsque des personnes se trouvent, en raison de l'offre médiatique ou de leurs propres choix, dans un espace médiatique fermé qui magnifie certains messages, par effet de répétition notamment, tout en les isolant des critiques et des points de vue différents des leurs⁴¹. La bulle épistémique, quant à elle, est le fruit du microciblage ainsi que de la personnalisation du classement des résultats de recherche en ligne et de l'offre sur le flux des médias sociaux. Cela fait en sorte que les personnes sont peu exposées à des contenus diversifiés qui pourraient contredire leurs opinions ou intérêts⁴².

Ces dynamiques ont déjà fait l'objet de plusieurs travaux en sciences sociales. Ces derniers ont mené plus récemment à la création de synthèses au sein de la littérature⁴³. Ces travaux s'intéressent surtout à l'influence des médias sociaux et numériques sur la qualité de la délibération publique, et non pas spécifiquement à l'impact de l'IA. Toutefois, les algorithmes de classement, de recommandation et de microciblage sont au cœur des processus qui génèrent des bulles épistémiques. À cela s'ajoute l'arrivée de robots conversationnels, qui permettent de diffuser automatiquement du contenu, un phénomène susceptible d'amplifier le processus de « magnification » des opinions qui donnent lieu aux chambres d'écho.

La polarisation mine la délibération de plusieurs manières. Tout d'abord, les clivages politiques profonds qu'elle provoque, et qui s'accompagnent parfois d'incivilités, rend difficile le maintien d'une conception publique et partagée du bien commun et de la justice permettant d'évaluer la légitimité des décisions politiques⁴⁴. De plus, une polarisation extrême peut saper les bases de la confiance sociétale nécessaire pour assurer la qualité des processus cognitifs.

³⁶ Cette idée a une longue histoire. Déjà, Aristote vantait les mérites de la délibération fondée sur la comparaison entre différents points de vue, mais c'est avec les premiers penseurs du libéralisme, comme Mill, et avec les premiers architectes de la constitution américaine (Hamilton) qu'elle devient intimement liée à l'idéal démocratique. Plusieurs penseurs politiques contemporains de la démocratie offrent un argument épistémique en faveur de la démocratie délibérative (Estlund [2009]; Landemore [2017]). Pour une critique, voir Brennan (2017).

³⁷ Landemore (2013).

³⁸ Arguedas et autres (2022).

³⁹ C. Sunstein (1999).

⁴⁰ *Ibid.*

⁴¹ Jamieson et Cappella (2010).

⁴² Pariser (2011).

⁴³ Arguedas et autres (2022); Costica Dumbrava (2021).

⁴⁴ Chambers et Kopstein (2001, 2021, 2022).



Les connaissances que nous avons ont une dimension sociale et dépassent l'horizon de notre expérience personnelle. C'est pourquoi nous devons nous en remettre aux témoignages de nos pairs qui ont observé des faits que nous n'avions pas nécessairement remarqués ainsi qu'aux avis des experts ayant acquis des connaissances spécialisées dans un domaine précis⁴⁵. Or, la polarisation causée par les chambres d'écho et les bulles épistémiques peut exacerber la méfiance envers les témoins et experts se situant « de l'autre côté » d'une question polarisante, au point où nous ne leur accordons plus la crédibilité qui devrait leur revenir.

2.2.2 Manipulation

Dans un contexte où il existe d'importantes inégalités de littératie⁴⁶, notamment numérique et algorithmique (p. ex. : inégalités relatives à la connaissance et à la compréhension du fonctionnement des médias numériques et des algorithmes), et où l'IA générative arrive facilement à créer des contenus pouvant duper les usagers⁴⁷, les risques de manipulation politique par la désinformation sont exacerbés⁴⁸. La manipulation est une forme d'influence insidieuse qui ne fait pas appel aux capacités réflexives des personnes. Elle fonctionne en partie parce qu'elle demeure cachée aux personnes ciblées⁴⁹. La désinformation dans les médias numériques alimentée par l'IA est une forme de manipulation puisqu'elle vise à modifier les croyances de manière insidieuse en exploitant les asymétries de connaissances relatives au fonctionnement de l'IA et des médias numériques⁵⁰.

Bien entendu, ce phénomène n'est pas nouveau. Depuis l'élection américaine de 2016 et la pandémie de COVID-19, des travaux ont été menés sur la désinformation et la dissémination de fausses nouvelles sur les plateformes numériques⁵¹. L'arrivée de l'IA générative qui produit facilement du texte, des images ainsi que des clips audio et vidéo propulse la désinformation à un niveau sans précédent⁵². L'usage de ce type d'IA est maintenant largement accessible, notamment depuis que des plateformes permettent de générer du contenu synthétique à partir de commandes exprimées dans des langues naturelles (français, anglais, etc.). Cela accentue fortement les impacts négatifs potentiels de la désinformation sur la délibération publique. Les gains en efficacité liés à l'usage de ces technologies augmentent l'efficacité et automatisent la création de tout type de contenus, ce qui ouvre la porte à la circulation d'un très haut volume de contenus synthétiques, notamment sur les médias sociaux, où la vitesse de propagation de l'information autant que de la désinformation ou de la mésinformation est très rapide, par exemple grâce aux robots conversationnels. Enfin, combinés à des techniques de microciblage, les hypertrucages deviennent d'autant plus convaincants qu'ils sont faits sur mesure pour susciter l'adhésion de publics cibles à un parti ou à une idéologie⁵³.

Les hypertrucages, ces créations visuelles et audio synthétiques prétendant représenter des personnes et des événements réels⁵⁴, suscitent un degré particulièrement élevé d'inquiétude chez les observateurs⁵⁵. D'une part, ces hypertrucages sont très difficiles à distinguer d'enregistrements authentiques, tant pour les êtres humains⁵⁶ que pour les algorithmes de détection⁵⁷. D'autre part, notre faible capacité à reconnaître ces hypertrucages remet en

⁴⁵ Christiano (2021, 2022).

⁴⁶ Rappelons qu'un Québécois sur deux n'atteint pas le niveau 3 de littératie (Langlois [2023]).

⁴⁷ Spital et autres (2023).

⁴⁸ Christiano (2022); Coeckelbergh (2022).

⁴⁹ Christiano (2022); C. R. Sunstein (2016).

⁵⁰ Christiano (2022).

⁵¹ Brown (2021); Chambers et Kopstein (2022); Costica Dumbrava (2021).

⁵² Buchanan et autres (2021); Goldstein et autres (2023); Paris et Donovan (2019).

⁵³ Cohen et Fung (2021).

⁵⁴ Helmus (2022).

⁵⁵ Chambers et Kopstein (2022); Citron et Chesney (2019); Goldstein et autres (2023); Helmus (2022); Horvitz, (2022); Paris et Donovan (2019); Pawelec (2022); Small et autres (2023).

⁵⁶ Köbis et autres (2021).

⁵⁷ Thompson et Hsu (2023).



question la confiance que nous avons dans les sources audio, vidéo et photographiques, normalement considérées comme des éléments solides de preuves empiriques permettant de vérifier les faits⁵⁸.

Il faut ajouter à cela que cette technologie et sa grande disponibilité suscitent de l'inquiétude parce qu'elles arrivent à un moment de notre histoire caractérisé par la montée des opérations d'influences politiques secrètes d'acteurs domestiques ou étrangers, privés ou publics, qui sont hostiles à la démocratie⁵⁹. Certains commentateurs soulignent par ailleurs que le risque le plus sérieux relativement aux hypertrucages ne réside pas tant dans la production par l'IA de fausses nouvelles ponctuelles ou à la pièce, mais plutôt dans la production d'« histoires synthétiques ». Le danger est que ces histoires synthétiques et ces fausses nouvelles, par leur intertextualité, se reprennent et se répondent mutuellement, formant ainsi de faux récits à partir desquels on peut bâtir un discours de propagande cohérent et réaliste⁶⁰.

La désinformation n'est pas la seule forme de manipulation qui s'insinue dans la délibération publique. Pour certains, le microciblage et les contenus adaptés aux profils personnels peuvent être si efficaces qu'ils constituent une forme d'« hypernudging », notamment lorsque le ciblage est constamment réactualisé selon l'activité des usagers d'outils numériques⁶¹. La notion de « nudge », issue de l'économie comportementale, réfère à des interventions sur l'architecture des choix qui s'offrent aux personnes, c'est-à-dire la manière dont différentes options leur sont présentées. Ces interventions visent à orienter la conduite des citoyens sans recourir à la coercition et sans imposer de pénalités⁶².

Par exemple, de la même manière qu'un supermarché place certains items pour favoriser l'achat impulsif, par exemple à une certaine hauteur sur les étagères ou près des caisses, une entreprise peut organiser le contenu de son site Web de sorte à favoriser un comportement, par ne pas quitter le site avant d'y avoir acheté un produit. De telles interventions sur l'architecture des choix sont problématiques lorsqu'elles sont insidieuses et qu'elles opèrent à l'insu des personnes; on peut alors affirmer qu'il s'agit de manipulation. Or, ce n'est pas nécessairement le cas de toutes les interventions de type *nudging*, en particulier celles qui visent des fins raisonnables de manière transparente⁶³.

Ainsi, l'enjeu, si l'on souhaite notamment réguler cette pratique, réside dans la difficulté à tracer la ligne entre le *nudging* « justifié » et la manipulation, puisqu'il s'agit de concepts gradués, c'est-à-dire présentant différents degrés⁶⁴. Par exemple, des chercheurs ont récemment démontré que la rédaction d'opinions politiques assistée par de grands modèles de langage comme ChatGPT modifie les opinions et attitudes des usagers d'une manière qui s'apparente à de la « persuasion latente »⁶⁵. Une telle rhétorique oriente, de manière délibérée ou non, les propos exprimés par les usagers qui sont exposés à des formulations et interventions particulières proposées par les grands modèles de langage. On peut se demander sous quelles conditions une telle influence relève de la manipulation ou non.

2.3 IA, démocratie et contestation

La délibération publique en démocratie a également pour fonction de permettre la contestation, par les citoyens, du pouvoir et des discours qui visent à asseoir sa légitimité.

⁵⁸ Paris et Donovan (2019).

⁵⁹ Chambers et Kopstein (2022); Goldstein et autres (2023); Risse (2021).

⁶⁰ Horvitz (2022).

⁶¹ Christiano (2022).

⁶² Thaler et Sunstein (2008).

⁶³ C. Sunstein (2015).

⁶⁴ C. R. Sunstein (2016).

⁶⁵ Jakesch et autres (2023).



D'une part, puisque chaque citoyen ne peut pas directement prendre part à l'élaboration des lois et des politiques publiques, il importe de mettre à leur disposition des moyens pour exprimer leur désaccord lorsque les lois et les grandes orientations d'un gouvernement leur semblent contraires au bien commun ou à leurs intérêts⁶⁶.

D'autre part, la contestation du pouvoir a besoin de s'incarner dans des contre-discours. Ces derniers s'inscrivent dans un rapport d'opposition aux puissances publiques et aux majorités électorales, ce qui introduit une dose d'adversité dans la joute politique⁶⁷. Il importe donc que cette contestation puisse se faire dans les termes, à l'intérieur des conceptions politiques et juridiques, et selon les formes de communication que privilégient les groupes sociaux, culturels, linguistiques et religieux marginalisés de même que par les autres institutions et organisations de la société civile (universités, syndicats, églises, etc.)⁶⁸.

La sphère médiatique numérique mobilise diverses capacités de l'IA. Cela peut entraver la contestation légitime du pouvoir et l'expression de contre-discours dans la sphère publique. D'un côté, étant donné que les groupes sociaux minoritaires sont moins représentés dans les données d'entraînement des algorithmes d'apprentissage automatique, il y a un risque que les algorithmes de classement, de recommandation et de création de contenus reflètent moins bien les points de vue de ces groupes. De l'autre, les technologies de reconnaissance faciale et de microciblage peuvent être utilisées à mauvais escient pour surveiller et réprimer certaines formes de contestations légitimes, mais dérangeantes aux yeux du gouvernement en place⁶⁹.

Enfin, la dimension contestataire de la délibération publique et le pluralisme politique compliquent la tâche visant à tracer la ligne entre la « mauvaise polarisation », soit celle qui érode la démocratie, et la « polarisation légitime », soit celle qui découle de la partialité et de l'adversité légitimes essentielles à la délibération⁷⁰.

2.4 Promesses de l'IA pour la délibération

Plusieurs travaux avancent que l'IA pourrait augmenter la qualité de la délibération publique. Tout d'abord, malgré leur opacité, les algorithmes peuvent contribuer à rendre l'exercice du pouvoir plus transparent, notamment en aidant à rendre certaines informations clés disponibles au public. Par exemple, les audits financiers et le contrôle des dépenses publiques peuvent exiger de traiter un volume de données trop vaste pour des organisations humaines. Dans ce contexte, l'IA peut amener le contrôle des dépenses à une échelle plus large en aidant les vérificateurs à traiter ces larges volumes de données de sorte à détecter plus d'irrégularités⁷¹.

L'IA peut également bonifier ou aider à assurer la qualité épistémique de la délibération. Certains avancent entre autres que l'IA peut nous aider à réaliser plus concrètement l'idéal délibératif à grande échelle⁷².

D'une part, l'IA peut délester les humains de certaines tâches coûteuses en temps, comme la modération, la traduction ou la vérification des faits (*fact-checking*). Les systèmes d'IA peuvent également analyser de vastes quantités de données composées de millions d'interventions dans la délibération publique numérique. Ainsi, ils peuvent cartographier les débats par le biais de techniques de visualisation de données, ce qui aide à situer sa propre position dans l'espace politique et à cerner les zones de consensus possibles, et d'analyse du langage permettant d'identifier des regroupements d'arguments similaires. Des robots conversationnels peuvent également semer des idées et arguments dans certaines assemblées et certains forums pour les diversifier. De plus, en quantifiant certains indicateurs de délibération, des agents conversationnels pourraient contribuer à encadrer une forme de « délibération démocratique augmentée »⁷³.

⁶⁶ Pettit (2000).

⁶⁷ Mouffe (2017); Tully (2008).

⁶⁸ Levy (2015); Young (2002).

⁶⁹ Robertson et autres (2020).

⁷⁰ Bonotti (2017).

⁷¹ Savaget et autres (2018).

⁷² Landemore (2023); Velikanov (2023).

⁷³ Hadfi et Ito (2022).



D'autres travaux avancent que plusieurs gains pour la délibération – taux de participation, qualité de l'engagement, inclusivité – pourraient être atteints par l'entremise de la création de plateformes numériques « ludifiées », c'est-à-dire du contenu qui emprunte certaines caractéristiques formelles des jeux (p. ex. : objectif commun, règles, indicateurs de progrès et autres caractéristiques attrayantes).⁷⁴ Enfin, on commence à observer certains résultats encourageants, mais très imparfaits, en ce qui concerne les algorithmes de détection des hypertrucages⁷⁵ et des robots conversationnels de lutte à la désinformation⁷⁶. Ces derniers ont été utilisés par exemple en France durant la pandémie de COVID-19⁷⁷.

3. Encadrement du rôle de l'IA dans le processus démocratique

Plusieurs États, dont le Royaume-Uni, les États-Unis et l'Union européenne (UE), ont déjà mis en place des mesures d'encadrement des systèmes d'IA ayant un impact sur le processus démocratique⁷⁸.

Les efforts de régulation visant à imposer des obligations et à faire valoir certaines interdictions oscillent en général entre deux approches. La première consiste en la mise en place d'obligations de divulgation et de transparence relatives au fait qu'une IA interagit avec des humains et aux données collectées. La seconde consiste en l'interdiction de certaines pratiques.

Par exemple, par rapport aux hypertrucages, certains États exigent maintenant que les développeurs de contenus synthétiques divulguent qu'ils ont été générés ou manipulés par l'IA. Ainsi, l'article 52 (3) de la *Loi sur l'IA* de l'UE stipule que « les utilisateurs d'un système d'intelligence artificielle qui génère ou manipule des images, des contenus audio ou vidéo qui ressemblent sensiblement à des personnes, des objets, des lieux ou d'autres entités ou événements existants et qui sembleraient faussement authentiques ou véridiques (*deepfakes*), doivent divulguer que le contenu a été généré ou manipulé artificiellement. »

D'autres législations ou propositions législatives, comme le *Malicious Deep Fake Prohibition Act* et le *Deep Fakes Accountability Act* aux États-Unis, interdisent certains usages des contenus synthétiques. Cela s'applique, par exemple, lorsque ceux-ci sont non consentis par les personnes qui y sont représentées, comme le suggèrent aussi certaines propositions d'amendement à la loi de l'UE, ou lorsqu'ils causent du tort.

Le même type de question se pose par rapport à la régulation du discours des robots conversationnels. Certains avancent que les États peuvent réguler ces discours de sorte à en interdire certains⁷⁹. D'autres soutiennent qu'on doit se limiter à une obligation de transparence, pour assurer aux usagers qu'ils soient avertis lorsqu'ils conversent avec une IA. De plus, cela permettrait de faire en sorte que le discours politique des robots conversationnels fasse l'objet des mêmes protections que les formes d'expression humaine, peu importe que les robots conversationnels expriment de manière autonome leurs propres vues ou celles d'êtres humains⁸⁰.

Par ailleurs, en plus de la régulation, les États peuvent mettre en place différentes politiques publiques visant à promouvoir des usages responsables de l'IA dans les campagnes électorales et la délibération publique. Les gouvernements peuvent, d'une part, agir à l'intérieur des plateformes numériques existantes pour mieux communiquer certains messages. Cela pourrait se faire notamment par l'entremise de robots conversationnels de lutte à la désinformation ou avec des modèles de langages qui sont entraînés avec des données orientant

⁷⁴ Gastil et Broghammer (2021).

⁷⁵ Thompson et Hsu (2023).

⁷⁶ Wang et Tanes-Ehle (2023).

⁷⁷ Raoul (2020).

⁷⁸ Holistic AI (2023).

⁷⁹ Fung et Lessig (2023); Lessig (2022).

⁸⁰ Weaver (2018).



délibérément leur « personnalité » politique. Ainsi, différents robots conversationnels promoutraient différentes positions politiques, ce qui augmenterait l'exposition des usagers à divers points de vue⁸¹.

Les États agissent également parfois sur l'environnement numérique lui-même en mettant en place des plateformes spécifiquement conçues pour produire des résultats bénéfiques sur le plan de la démocratie, plutôt que pour maximiser les revenus publicitaires. C'est notamment ce que font certaines administrations qui, à l'instar de la plateforme vTaiwan, utilisent un système comme Pol.is pour encadrer la délibération publique. Le site Web de Pol.is recense plusieurs cas d'utilisation de systèmes par différents gouvernements, surtout locaux pour l'instant⁸². Enfin, les États peuvent également agir sur les compétences et dispositions des citoyens par le biais de diverses initiatives visant à promouvoir la littératie numérique et algorithmique, à l'instar de l'Initiative de citoyenneté numérique du Canada⁸³.

⁸¹ Knight (2023); Wong (2023).

⁸² <https://compdemocracy.org/Case-studies/>.

⁸³ Oeldorf-Hirsch et Neubaum (2023); Polanco-Levican et Salvo-Garrido (2022).



Bibliographie de la section Démographie et intelligence artificielle

Arguedas, A. R., Craig T. Robertson, Richard Fletcher, et Rasmus Kleis Nielsen. (2022). *Echo chambers, filter bubbles, and polarisation : A literature review*. Reuters Institute. <https://reutersinstitute.politics.ox.ac.uk/echo-chambers-filter-bubbles-and-polarisation-literature-review>.

Beckman, L., Hultin Rosenberg, J., et Jebari, K. (2022). *Artificial intelligence and democratic legitimacy. The problem of publicity in public authority*. AI & Society. <https://doi.org/10.1007/s00146-022-01493-0>.

Bonotti, M. (2017). *Partisanship and Political Liberalism in Diverse Societies*. Oxford University Press.

Borgesius, F. J. Z., Möller, J., Kruikemeier, S., Fathaigh, R. Ó., Irion, K., Dobber, T., Bodo, B., et Vreese, C. de. (2018). *Online Political Microtargeting: Promises and Threats for Democracy*. Utrecht Law Review. (Nº 1), 14(1), 1. <https://doi.org/10.18352/ulr.420>.

Brennan, J. (2017). *Against Democracy*. Princeton University Press. <https://press.princeton.edu/books/paperback/9780691178493/against-democracy>.

Brown, É. (2021). *Regulating the spread of online misinformation*. In *The Routledge Handbook of Political Epistemology*. Routledge.

Buchanan, B., Lohn, A., Musser, M., et Sedova, K. (2021). *Truth, Lies, and Automation*. Centre for Security and Emerging Technology. <https://cset.georgetown.edu/publication/truth-lies-and-automation/>.

Caron, S. (2023). *Élections à l'ère des mégadonnées : les conséquences de l'hyperclientélisme politique au Canada*. Mémoire de Maîtrise. Université du Québec à Montréal. <https://archipel.uqam.ca/16514/>.

Chambers, S., et Kopstein, J. (2001). *Bad Civil Society*. Political Theory, 29(6), 837-865.

Chambers, S., et Kopstein, J. S. (4 mai 2021). Revisiting "Bad Civil Society". HistPhil. <https://histphil.org/2021/05/04/revisiting-bad-civil-society/>.

Chambers, S., et Kopstein, J. (2022). *Wrecking the public sphere: The new authoritarians' digital attack on pluralism and truth*. Constellations.

Chomanski, B. (2022). *Legitimacy and automated decisions: the moral limits of algocracy*. Ethics and Information Technology, 24(3), 34. <https://doi.org/10.1007/s10676-022-09647-w>.

Christiano, T. (2010). *The Constitution of Equality: Democratic Authority and Its Limits*. Oxford University Press.

Christiano, T. (2021). *The Basis of Political Equality*. In E. Edenberg et M. Hannon (Éds.), *Political Epistemology* (p. 114-134). Oxford University Press. <https://doi.org/10.1093/oso/9780192893338.003.0008>.

Christiano, T. (2022). *Algorithms, Manipulation, and Democracy*. Canadian Journal of Philosophy, 52(1), 109-124. <https://doi.org/10.1017/can.2021.29>.

Citron, D., et Chesney, R. (2019). *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. California Law Review, 107(6), 1753.

Coeckelbergh, M. (2022). *Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence*. AI and Ethics, 1-10. <https://doi.org/10.1007/s43681-022-00239-4>.



Cohen, J., et Fung, A. (2021). 1. *Democracy and the Digital Public Sphere*. In 1. *Democracy and the Digital Public Sphere* (p. 23-61). University of Chicago Press. <https://doi.org/10.7208/9780226748603-002>.

Costica Dumbrava. (2021). *Les principaux risques des médias sociaux pour la démocratie : Risques liés à la surveillance, à la personnalisation, à la désinformation, à la modération et au microciblage*. Parlement européen. [https://www.europarl.europa.eu/thinktank/fr/document/EPRS_IDA\(2021\)698845](https://www.europarl.europa.eu/thinktank/fr/document/EPRS_IDA(2021)698845).

Danaher, J. (2016). *The Threat of Algocracy: Reality, Resistance and Accommodation*. *Philosophy et Technology*, 29(3), 245-268. <https://doi.org/10.1007/s13347-015-0211-1>.

De Fine Licht, K., et de Fine Licht. (2020). *Artificial intelligence, transparency, and public decision-making*. *AI & Society*, 35(4), 917-926. <https://doi.org/10.1007/s00146-020-00960-w>.

Duberry, J. (2022a). *Artificial Intelligence and Democracy*. Edward Elgar Publishing. <https://redshelf.com/app/ecom/book/2127790/artificial-intelligence-and-democracy-2127790-9781788977319-jerome-duberry>.

Duberry, J. (2022b). *Artificial Intelligence and Democracy: Risks and Promises of AI-Mediated Citizen-Government Relations*. In *Artificial Intelligence and Democracy*. Edward Elgar Publishing. <https://www.elgaronline.com/monobook-oa/book/9781788977319/9781788977319.xml>.

Estlund, D. (2009). *Democratic Authority: A Philosophical Framework*. In *Democratic Authority*. Princeton University Press. <https://doi.org/10.1515/9781400831548>.

Folco, J. D., et Martineau, J. (2021). *Cartographier les résistances à l'ère du capital algorithmique*. *Revue Possibles*, 45 (1), 1.

Fung, A., et Lessig, L. (2023). *How AI Could Take Over Elections—And Undermine Democracy*. *Scientific American*. <https://www.scientificamerican.com/article/how-ai-could-take-over-elections-and-undermine-democracy/>.

Gastil, J., et Broghammer, M. (2021). *Linking Theories of Motivation, Game Mechanics, and Public Deliberation to Design an Online System for Participatory Budgeting*. *Political Studies*, 39(1), 7-25.

Goldstein, J. A., DiResta, R., Sastry, G., Musser, M., Gentzel, M., et Sedova, K. (2023). *Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations*. <https://fsi.stanford.edu/publication/generative-language-models-and-automated-influence-operations-emerging-threats-and>.

Gorton, W. A. (2016). *Manipulating Citizens: How Political Campaigns' Use of Behavioral Social Science Harms Democracy*. *New Political Science*, 38(1), 61-80. <https://doi.org/10.1080/07393148.2015.1125119>.

Gutmann, A., et Thompson, D. (2004). *Why Deliberative Democracy?* Princeton University Press. <https://press.princeton.edu/books/paperback/9780691120195/why-deliberative-democracy>.

Habermas, J. (1987). *La théorie de l'agir communicationnel. Tome 1: Rationnalité de l'agir et rationalisation de la société*. Fayard.

Hadfi, R., et Ito, T. (2022). *Augmented Democratic Deliberation: Can Conversational Agents Boost Deliberation in Social Media?* Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems, 1794-1798.



Helmus, T. C. (2022). *Artificial Intelligence, Deepfakes, and Disinformation: A Primer*. RAND Corporation. <https://www.rand.org/pubs/perspectives/PEA1043-1.html>.

Hidalgo, C. (2019). *A bold idea to replace politicians*. TED Talk. https://www.ted.com/talks/cesar_hidalgo_a_bold_idea_to_replace_politicians.

Himmelreich, J. (2022). *Should We Automate Democracy?* In, *The Oxford Handbook of Digital Ethics* C. Véliz (Éd.) (p. 0). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198857815.013.33>.

Holistic AI. (2023). *The State of AI Regulations in 2023*. Holistic AI. <https://www.holisticai.com/papers/the-state-of-ai-regulations-in-2023>.

Horton, C. (2018). *The simple but ingenious system Taiwan uses to crowdsource its laws*. MIT Technology Review. <https://www.technologyreview.com/2018/08/21/240284/the-simple-but-ingenious-system-taiwan-uses-to-crowdsource-its-laws/>.

Horvitz, E. (2022). *On the Horizon : Interactive and Compositional Deepfakes*. International Conference on Multimodal Interaction, 653-661. <https://doi.org/10.1145/3536221.3558175>.

Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L., et Naaman, M. (2023). *Co-Writing with Opinionated Language Models Affects Users' Views*. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1-15. <https://doi.org/10.1145/3544548.3581196>.

Jamieson, K. H., et Cappella, J. N. (2010). *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford University Press.

Jones, D., Snider, C., Nassehi, A., Yon, J., et Hicks, B. (2020). *Characterising the Digital Twin: A systematic literature review*. CIRP Journal of Manufacturing Science and Technology, 29, 36-52. <https://doi.org/10.1016/j.cirpj.2020.02.002>.

Knight, W. (24 juillet 2023). *Meet ChatGPT's Right-Wing Alter Ego*. Wired. <https://www.wired.com/story/fast-forward-meet-chatgpts-right-wing-alter-ego/>.

Köbis, N. C., Doležalová, B., et Soraperra, I. (2021). *Fooled twice: People cannot detect deepfakes but think they can*. IScience, 24(11), 103364. <https://doi.org/10.1016/j.isci.2021.103364>.

Kreiss, D. (2017). *Micro-targeting, the quantified persuasion*. Internet Policy Review, 6(4). <https://policyreview.info/articles/analysis/micro-targeting-quantified-persuasion>.

Landemore, H. (2013). *Democratic reason: politics, collective intelligence, and the rule of the many*. Princeton University Press.

Landemore, H. (2017). *Democratic Reason*. Princeton University Press. <https://press.princeton.edu/books/paperback/9780691176390/democratic-reason>.

Landemore, H. (2023). *Can AI bring deliberative democracy to the masses?* The Living Library. <https://thelivinglib.org/can-ai-bring-deliberative-democracy-to-the-masses/>.

Langlois, P. (24 janvier 2023). *Estimation des indices de littératie par MRC*. Fondation pour l'alphabétisation. <https://www.treaq.ca/recherche-et-publications/projection-de-lindice-de-litteratie-au-quebec-en-2022-un-progres-qui-met-en-lumiere-un-enjeu-importent/>.



Lavigne, M. (2022). *Microtargeting*. In *Elgar Encyclopedia of Technology and Politics* (p. 231-235). Edward Elgar Publishing. <https://www.elgaronline.com/edcollchap/book/9781800374263/b-9781800374263.microtargeting.lavigne.xml>.

Lessig, L. (2022). *The First Amendment Does Not Protect Replicants*. In *Social Media, Freedom of Speech, and the Future of our Democracy*, L. C. Bollinger et G. R. Stone (Éds.), (p. 0). Oxford University Press. <https://doi.org/10.1093/oso/9780197621080.003.0016>.

Lever, A. (2012). *On Privacy*. Routledge. <https://www.routledge.com/On-Privacy/Lever/p/book/9780415395700>.

Levy, J. T. (2015). *Rationalism, Pluralism, and Freedom*. Oxford University Press.

Maclure, J. (2021). *AI, Explainability and Public Reason: The Argument From the Limitations of the Human Mind*. *Minds and Machines*, 31(3), 421-438. <https://doi.org/10.1007/s11023-021-09570-x>.

Maclure, J., et Saint-Pierre, M.-N. (2018). *Le nouvel âge de l'intelligence artificielle : Une synthèse des enjeux éthiques*. *Les Cahiers de propriété intellectuelle*, 30 (3), 741-766.

Mainz, J. T., Sønderholm, J., et Uhrenfeldt, R. (2022). *Artificial intelligence and the secret ballot*. *AI & Society*. <https://doi.org/10.1007/s00146-022-01551-7>.

Mattu, J. A., Jeff Larson, Lauren Kirchner, Surya. (2016). *Machine Bias*. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Miller, C. (2019). *Taiwan is making democracy work again. It's time we paid attention*. Wired. <https://www.wired.co.uk/article/taiwan-democracy-social-media>.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., et Floridi, L. (2016). *The ethics of algorithms: Mapping the debate*. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>.

Mouffe, C. (2017). *Le Paradoxe démocratique*. Archives de la critique d'art. Paris Beaux-Arts.

Oeldorf-Hirsch, A., et Neubaum, G. (2023). *What do we know about algorithmic literacy? The status quo and a research agenda for a growing field*. *New Media & Society*. <https://doi.org/10.1177/14614448231182662>.

Ouellet, M. (2021). *Pour une théorie critique de la gouvernance algorithmique et de l'intelligence artificielle*. *Revue tic&société*, Vol. 15, N° 1 | 1er semestre 2021. <https://doi.org/10.4000/ticsociete.5603>.

Paris, B., et Donovan, J. (18 septembre 2019). *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence*. Data & Society. Data & Society Research Institute. <https://datasociety.net/library/deepfakes-and-cheap-fakes/>.

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin.

Pawelec, M. (2022). *Deepfakes and Democracy (Theory) : How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions*. *Digital Society: Ethics, Socio-Legal and Governance of Digital Technology*, 1(2), 19. <https://doi.org/10.1007/s44206-022-00010-6>.

Pettit, P. (2000). *Democracy, Electoral and Contestatory*. Nomos, 42, 105-144.



Polanco-Levicán, K., et Salvo-Garrido, S. (2022). *Understanding Social Media Literacy: A Systematic Review of the Concept and Its Competences*. IJERPH, 19(14), 1-16.

Raoul, G. (2020). *Covid-BOT : Un chatbot open source français pour lutter contre la désinformation*. LeMagIT. <https://www.lemagit.fr/actualites/252480549/Coronavirus-le-chatbot-francais-Covid-Bot-lutte-contre-la-desinformation>.

Rawls, J. (2005). *Political Liberalism: Expanded Edition*. Columbia University Press.

Rawls, J. (2008). *La justice comme équité*. La découverte. https://www.editionsladecouverte.fr/la_justice_comme_equite-9782707154606.

Reiman, J. (1995). *Driving to the Panopticon : A Philosophical Exploration of the Risks to Privacy Posed by the Highway Technology of the Future*. Santa Clara High Technology Law Journal, 11(1), 27.

Risse, M. (2021). *Artificial Intelligence and the Past, Present, and Future of Democracy*. Carr Centre Discussion Paper Series.

Robertson, K., Khoo, C., et Song, Y. (2020). *To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada*. Citizen Lab, University of Toronto.

<https://citizenlab.ca/2020/09/to-surveil-and-predict-a-human-rights-analysis-of-algorithmic-policing-in-canada/>.

Rosanvallon, P. (2011). *La Société des égaux*. Seuil.

Sanders, N. E. (2023). *AI Will Upend Election Season*. The Atlantic. <https://www.theatlantic.com/technology/archive/2023/04/ai-generated-political-ads-election-candidate-voter-interaction-transparency/673893/>.

Savaget, P., Chiarini, T., et Evans, S. (2018). *Empowering political participation through artificial intelligence*. Science & Public Policy, 46(3), 369-380. <https://doi.org/10.1093/scipol/scy064>.

Schumpeter, J. A. (2008). *Capitalism, Socialism, and Democracy: Third Edition*. Harper Perennial Modern Classics.

Scola, N. (20 novembre 2016). *How chatbots are colonizing politics*. Politico. <https://www.politico.com/story/2016/10/chatbots-are-invading-politics-229598>.

Shoshana Zuboff. (2020). *L'Âge du capitalisme de surveillance*. Éditions Zulma.

Simons, J. (2023). *Algorithms for the People. Democracy in the Age of AI*. Princeton University Press.

Small, C. T., Vendrov, I., Durmus, E., Homaei, H., Barry, E., Cornebise, J., Suzman, T., Ganguli, D., et Megill, C. (2023). *Opportunities and Risks of LLMs for Scalable Deliberation with Polis*. arXiv:2306.11932. <https://www.haoshuo.com/post/64949a5873041147815d4a9a>.

Spitale, G., Biller-Andorno, N., et Germani, F. (2023). *AI model GPT-3 (dis)informs us better than humans*. Science Advances, 9(26), eadh1850. <https://doi.org/10.1126/sciadv.adh1850>.

Sunstein, C. (1999). *The Law of Group Polarization. Coase-Sandor Working Paper Series in Law and Economics*. Chicago Unbound. https://chicagounbound.uchicago.edu/law_and_economics/542.



Sunstein, C. (2015). *The Ethics of Nudging*. Yale Journal on Regulation.
<https://openyls.law.yale.edu/handle/20.500.13051/8225>.

Sunstein, C. R. (Éd.). (2016). *Fifty Shades of Manipulation*. In *The Ethics of Influence: Government in the Age of Behavioral Science* (p. 78-115). Cambridge University Press.
<https://doi.org/10.1017/CBO9781316493021.005>.

Tappin, B. M., Wittenberg, C., Hewitt, L. B., Berinsky, A. J., et Rand, D. G. (2023). *Quantifying the potential persuasive returns to political microtargeting*. Proceedings of the National Academy of Sciences, 120(25), e2216261120.
<https://doi.org/10.1073/pnas.2216261120>.

Thaler, R., et Sunstein, C. R. (2008). *Nudge*. Penguin Random house.
<https://www.penguinrandomhouse.com/books/690485/nudge-by-richard-h-thaler-and-cass-r-sunstein/>.

Thiel, T. (2022). *Artificial Intelligence and Democracy*. Henrich Böll Stiftung.
<https://il.boell.org/en/2022/01/06/artificial-intelligence-and-democracy>.

Thompson, S. A., et Hsu, T. (28 juin 2023). *How Easy Is It to Fool A.I.-Detection Tools?* The New York Times.
<https://www.nytimes.com/interactive/2023/06/28/technology/ai-detection-midjourney-stable-diffusion-dalle.html>.

Tully, J. (2008). *Public Philosophy in a New Key*. Cambridge University Press.
<https://www.cambridge.org/core/books/public-philosophy-in-a-new-key/EE0BCF3CE3DEF04603E41E3D4E9CEDA4>.

Velikanov, C. (2023). *Mass Online Deliberation*. https://www.academia.edu/12031548/Mass_Online_Deliberation.

Wang, J., et Tanes-Ehle, Z. (2023). *Examining the Effects of Conversational Chatbots on Changing Conspiracy Beliefs about Science: the Paradox of Interactivity*. Journal of Broadcasting & Electronic Media, 67(1), 68-89.
<https://doi.org/10.1080/08838151.2022.2153842>.

Weaver, J. F. (16 janvier 2018). *Why Robots Deserve Free Speech Rights*. Slate.
<https://slate.com/technology/2018/01/robots-deserve-a-first-amendment-right-to-free-speech.html>.

Wong, M. (2023, mars 2). *Conspiracy Theories Have a New Best Friend*. The Atlantic.
<https://www.theatlantic.com/technology/archive/2023/03/generative-ai-disinformation-synthetic-media-history/673260/>.

Young, I. M. (2002). *Inclusive Political Communication*. In *Inclusion and Democracy*, I. M. Young (Éd.), (p. 0). Oxford University Press. <https://doi.org/10.1093/0198297556.003.0003>.



II. Impact environnemental et intelligence artificielle

Grâce à la capacité grandissante de l'IA à traiter de vastes quantités de données, on dénote un intérêt de plus en plus important envers son utilisation dans la lutte contre les changements climatiques et l'atténuation de leurs effets. Lorsqu'il est question d'environnement, il faut agir rapidement et jouer avec plusieurs leviers. L'IA est l'un de ces leviers⁸⁴.

Le chercheur David Rolnick ainsi que plusieurs autres experts dans le domaine soulignent toutefois que l'IA n'est pas une solution miracle pour résoudre le problème complexe des changements climatiques; elle n'est qu'un outil parmi d'autres⁸⁵. Bien qu'elle ait démontré des résultats positifs dans certains secteurs, l'IA a également des conséquences néfastes avérées sur l'environnement. Il faut savoir évaluer et mitiger ces impacts⁸⁶.

Nous proposons de les survoler, qu'ils soient directs ou indirects, positifs ou négatifs, pour ensuite s'attarder au rôle que peuvent jouer les normes internationales dans le développement et le déploiement de l'IA. Finalement, nous présenterons certaines initiatives hors Québec qui font l'objet de discussions sur la scène internationale. Ces initiatives portent sur l'encadrement de l'IA dans le contexte de la lutte contre les changements climatiques et de la protection de l'environnement.

1. Conséquences environnementales de l'IA

Il existe plusieurs méthodes pour évaluer les impacts environnementaux de l'IA. Les principaux cadres d'analyse dans la littérature tendent à les regrouper sous deux grandes catégories, soit les impacts directs et indirects des SIA sur l'environnement. Les conséquences directes découlent de l'existence matérielle de ces systèmes et renvoient au cycle de vie d'un SIA : production, transport, opération et fin de vie. Les conséquences indirectes réfèrent plutôt aux différentes utilisations de l'IA.

Alors que les impacts directs des SIA sur l'environnement sont reconnus comme majoritairement négatifs, les impacts indirects peuvent être ou bien positifs, ou bien négatifs. Les sections suivantes recensent quelques-uns de ces impacts et les enjeux qui y sont liés. Cependant, il ne s'agit pas d'une analyse exhaustive des impacts des SIA sur l'environnement.

Impacts directs

Les impacts directs des SIA sont largement négatifs. Ils sont associés au cycle de vie des SIA et de leurs infrastructures physiques, ces dernières étant liées à la capacité de calcul nécessaire à l'IA (*AI compute*).

La norme ISO14040-44 inclut les étapes suivantes au cycle de vie d'un produit ou d'un service: (1) production; (2) transport; (3) opération; (4) fin de vie⁸⁷. Il existe également plusieurs cadres d'évaluation des impacts environnementaux plus spécifiques aux SIA.

⁸⁴ Dennis et autres (2023).

⁸⁵ Cowls et autres (2023); Rolnick et autres (2019).

⁸⁶ Dennis et autres (2023).

⁸⁷ OCDE (2022a).



À cet effet, l'OCDE a publié en 2022 le Cadre de classification⁸⁸ pour les SIA, qui fait état des différentes phases de leur cycle de vie suivantes :

- I. Planification et design;
- II. Collecte et traitement des données;
- III. Création et utilisation du modèle;
- IV. Vérification et validation du modèle;
- V. Déploiement;
- VI. Opération et surveillance du modèle.

La phase VI permet de calculer les impacts de toutes les autres. Les impacts peuvent être regroupés en différentes catégories, par exemple : changement climatique résultant des émissions de gaz à effet de serre (GES); consommation non viable des ressources abiotiques; demande d'énergie cumulée; consommation d'eau.

Bien que l'exercice soit complexe, certaines études ont tenté de calculer l'empreinte environnementale des SIA. Pour ce faire, elles ont évalué la consommation des ressources (matériaux, énergie, etc.) et les émissions – incluant les GES – associées à chacune des étapes nécessaires pour entraîner un SIA. Il a donc fallu considérer le cycle de vie des ressources computationnelles nécessaires aux phases I, II, III et IV.

Par exemple, l'organisation Hugging Face a tenté de calculer les émissions de CO₂ de BLOOM, leur grand modèle de langage (*Large Language Model*), comportant 176 milliards de paramètres, et de le comparer aux performances environnementales d'autres modèles similaires sur le marché. Plusieurs facteurs ont été considérés : l'énergie utilisée pour entraîner le modèle sur un superordinateur; l'énergie nécessaire à la fabrication du matériel du superordinateur et à la maintenance de son infrastructure informatique; l'énergie utilisée pour faire fonctionner BLOOM.

Hugging Face a estimé que la phase d'entraînement de BLOOM avait généré 25 tonnes de CO₂, soit l'équivalent de 250 vols aller-retour entre Montréal et Toronto⁸⁹. Cependant, les chercheurs ont constaté que cette quantité doublait lorsqu'ils prenaient en compte les émissions produites par la fabrication du matériel informatique utilisé pour la formation de l'infrastructure informatique et l'énergie nécessaire pour faire fonctionner BLOOM après son entraînement. BLOOM étant entraîné sur un superordinateur français propulsé principalement par l'énergie nucléaire, son empreinte carbone n'est pas comparable à de grands modèles de langage similaires. Par exemple, selon les données disponibles, on estime que GPT-3 d'OpenAI et OPT de Meta émettent respectivement plus de 500 et 75 tonnes de dioxyde de carbone pendant la phase d'entraînement seulement. Cela s'explique en grande partie par la différence de l'intensité carbone de la source d'énergie utilisée pour l'entraînement d'un modèle⁹⁰. De plus, dans le cas des grands modèles de langages, c'est souvent l'usage qui exige une plus grande consommation d'électricité et qui produit le plus de CO₂⁹¹, puisque la phase d'inférence d'un SIA déployé à grande échelle entraîne généralement une utilisation ininterrompue⁹².

Ensuite, un autre aspect important concerne les impacts environnementaux générés par les centres de données qui sont le pilier du déploiement des SIA. Selon l'International Energy Agency (IEA): « Les centres de données et les réseaux de transmission de données qui sous-tendent la numérisation représentaient environ 300 mégatonnes d'équivalent en CO₂ en 2020 (y compris les émissions intrinsèques), soit 0,9 % des émissions de GES liées à l'énergie (ou 0,6 % des émissions totales de GES) »⁹³. Au Canada, les centres de données nécessitent 10 % de la totalité de

⁸⁸ OCDE (2022b).

⁸⁹ Lagueux-Beloin (2019).

⁹⁰ Luccioni et autres (2022).

⁹¹ Ludvigsen (2022).

⁹² Cowls et autres (2023).

⁹³ IEA (2022).



l'énergie consommée par le numérique et représentent 17 % de l'empreinte carbone totale du numérique. Au Québec, les centres de données représentent 17 % de l'énergie totale du numérique et 29 % de l'empreinte carbone totale du numérique⁹⁴. Évidemment, l'apprentissage automatique étant en forte croissance, les implications liées à la consommation énergétique et les émissions des centres de données seront plus grandes. Au cours des dernières années, chez Facebook (Meta), la demande de calcul pour l'entraînement de l'apprentissage automatique et de l'inférence, qui augmentent respectivement de 150 % et de 105 % par an, a dépassé l'utilisation globale d'énergie dans les centres de données qui augmente de 40 % annuellement. Google signale que l'apprentissage automatique ne représente que 10 à 15 % de sa consommation totale d'énergie, alors qu'il représente 70 à 80 % de la demande informatique globale⁹⁵.

Impacts indirects

Les impacts indirects de l'IA sur l'environnement découlent principalement de ses utilisations et rassemblent les possibilités avérées ainsi que les effets négatifs des SIA. Il est toutefois beaucoup plus difficile d'évaluer et de quantifier ce type d'impact puisque les effets sont plus diffus, comparativement aux impacts directs⁹⁶.

1.1.1 Retombées positives indirectes : occasions à saisir et utilisations

Les SIA présentent un potentiel avéré. Ils auraient même déjà démontré des bénéfices notables dans la lutte contre les changements climatiques⁹⁷. Les possibilités qu'offrent les SIA en matière de classification, de prédiction et même de prise de décisions permettent, à plus grande échelle, d'améliorer notre compréhension du problème et de faciliter une réponse efficace⁹⁸. La capacité des SIA à traiter de vastes quantités de données non structurées et multidimensionnelles, à l'aide de techniques d'optimisation sophistiquées, permet également de mieux comprendre les ensembles de données pluridimensionnelles et, ainsi, mieux prédire les tendances futures⁹⁹.

Dans une [étude](#), le chercheur Josh Cowls et ses collègues recensent plusieurs exemples où l'IA pourrait aider les populations à s'adapter aux changements climatiques et contribuer à bâtir des sociétés plus résilientes face à ces enjeux environnementaux. Cela passerait, entre autres, par une meilleure prédiction des phénomènes climatiques et océaniques¹⁰⁰. En effet, l'IA pourrait permettre une meilleure anticipation des phénomènes météorologiques violents – fortes pluies, feux de forêt, etc. – ainsi que les conséquences qui en découlent, comme les mouvements de migrations climatiques.

⁹⁴ The Shifters Montreal (2020).

⁹⁵ IEA (2022).

⁹⁶ OCDE (2022).

⁹⁷ Rolnick et autres (2019).

⁹⁸ Cowls et autres (2023).

⁹⁹ Huntingford et autres (2019).

¹⁰⁰ Ham et autres (2019).



Selon cette même étude, une lutte efficace contre les changements climatiques passe aussi par un éventail de solutions diversifiées. Par exemple, l'IA pourrait aider à améliorer l'efficacité énergétique dans certains secteurs industriels, ou encore à mieux comprendre l'impact carbone du béton dans le milieu de la construction. En matière d'actions pour l'environnement, le rapport [Harnessing Artificial Intelligence for the Earth](#), publié en 2018 par le Forum économique mondial (FEM), résume le potentiel de l'IA pour plus de 80 cas d'utilisation répartis en six « défis environnementaux », soit :

- Les changements climatiques;
- La préservation de la biodiversité;
- La santé des océans;
- La sécurité hydrique;
- La qualité de l'air;
- La résilience face aux conditions météorologiques et aux catastrophes naturelles.

Le FEM avance que les utilisations actuelles relèvent principalement de SIA automatisés et assistés servant à tirer parti des vastes ensembles de données que composent ces six grandes thématiques. Possiblement, les utilisations seraient plutôt propulsées par des SIA étant capables de prise de décisions autonome. Cela pourrait potentiellement créer de nouvelles occasions à saisir, mais également exacerber les risques environnementaux. D'où l'importance pour les différents acteurs de placer les considérations de durabilité au cœur de l'innovation, du développement et de l'utilisation de l'IA à plus grande échelle¹⁰¹.

L'IA démontre aussi un potentiel fort intéressant dans certains domaines, notamment celui de la préservation de la biodiversité. Actuellement, l'IA peut y contribuer de trois manières, soit :

- en améliorant la protection des espèces menacées et en voie d'extinction;
- en surveillant en temps réel les changements dans les grands écosystèmes;
- en modélisant et en anticipant plusieurs scénarios futurs et leurs impacts sur la biodiversité.

Premièrement, les technologies de reconnaissance d'images propulsées par les SIA se présentent comme un outil prometteur pour mieux protéger les espèces menacées et en voie de disparition. En effet, elles permettent de mieux comprendre leur dénombrement et de connaître les déplacements des espèces¹⁰². Deuxièmement, les outils de cartographie satellitaire et aérienne, ajoutés à la capacité d'analyse de l'IA, rendent possible la vue en temps réel des changements dans les grands écosystèmes, permettant la génération de données très précises sur les émissions de carbone et la pollution atmosphérique, l'aménagement du territoire, etc. Troisièmement, les capacités computationnelles de l'IA permettent de modéliser et d'anticiper plusieurs scénarios futurs et leurs impacts sur la biodiversité, comme les changements climatiques et les catastrophes naturelles, la propagation de virus ou encore les impacts de toute interaction humaine avec la biodiversité¹⁰³.

À cet effet, un [article](#)¹⁰⁴ publié dans la revue *Nature Sustainability* présente un cadre sur la priorisation de la conservation du patrimoine. Basé sur l'apprentissage par renforcement, ce cadre utilise des données simulées et empiriques. Avec un budget limité, le modèle parvient à des résultats nettement meilleurs que d'autres logiciels, atteignant les objectifs de conservation de manière plus fiable et générant des cartes de priorisation plus faciles à interpréter. Le modèle cité dans l'article explore également l'impact de la sensibilité d'espèces spécifiques aux perturbations locales qui varient géographiquement et aux changements climatiques, par exemple la construction d'une nouvelle route, la présence d'activités minières ou toute autre forme d'activités économiques ayant un impact

¹⁰¹ FEM (2018).

¹⁰² Norouzzadeh et autres (2018).

¹⁰³ Duberry (2021).

¹⁰⁴ Silvestro et autres (2022).



négatif sur les écosystèmes naturels. À long terme, l'IA pourrait également augmenter la transparence et la responsabilité des actions pour soutenir la protection de la biodiversité¹⁰⁵.

Dans le même ordre d'idées, la littérature reconnaît une certaine tendance à percevoir l'IA comme un outil pouvant contribuer à l'atteinte des [objectifs de développement durable \(ODD\) établis par les Nations unies](#), bien que les impacts comportent certaines limites¹⁰⁶. Dans les dernières années, plusieurs chercheurs ont tenté d'identifier et d'évaluer l'impact des SIA et d'évaluer pour quels ODD l'IA pourrait être un catalyseur¹⁰⁷. Un [article](#)¹⁰⁸ s'appuyant sur un processus critique basé sur le consensus d'experts fait une évaluation détaillée du rôle de l'IA dans l'atteinte de chacun des objectifs, à savoir si l'IA joue un rôle de catalyseur, ou plutôt un rôle négatif ou inhibiteur. L'analyse réalisée démontre que l'IA peut potentiellement agir comme catalyseur pour 134 objectifs, soit 79 % de tous les ODD établis par les Nations Unies. Les bénéfices apportés par l'IA dans l'atteinte de ces objectifs peuvent se traduire par la capacité d'analyser de grands ensembles de données interconnectées, pour mettre en place une action conjointe pour la préservation de l'environnement. À l'inverse, l'IA peut avoir des impacts négatifs sur 59 des objectifs, soit 35 % du total des ODD.

Cette analyse a été réalisée en séparant les objectifs selon trois principaux piliers : la société, l'économie et l'environnement. Si l'on s'attarde au pilier environnemental, il semblerait que l'IA pourrait contribuer à atteindre les cibles de 25 objectifs (93 %), alors qu'il pourrait avoir des impacts négatifs dans l'atteinte de 8 objectifs (30 %). En pratique, chaque interconnexion entre l'IA et les différents objectifs a été pondérée selon l'applicabilité et la pertinence de chacune des références pour évaluer une interconnexion spécifique et, éventuellement, trouver les lacunes de la recherche.

Bien que la prise en compte du type de preuve ait un effet relativement faible sur les impacts positifs – on constate une réduction des cibles affectées positivement, qui passent de 79 % à 71 % –, on observe toutefois une réduction plus importante, passant de 35 % à 23 %, des cibles à impact négatif de l'IA. Quant au pilier environnemental, avec la pondération pratique, il y a une diminution de 8 % (on passe de 93 % à 85 %) des cibles affectées positivement et une diminution de 18 % (on passe de 30 % à 12 %) des cibles affectées négativement. Cette importante diminution peut s'expliquer par le manque de preuves tangibles des impacts négatifs pour chacun des objectifs, malgré de fortes indications qui pointent en ce sens et qui ouvrent la porte du même coup à de futures recherches. Les possibilités d'utilisation ayant des impacts positifs sont toutefois nombreuses pour les objectifs 13 (mesures relatives à la lutte contre les changements climatiques), 14 (vie aquatique) et 15 (vie terrestre). Par contre, les auteurs mettent en garde contre la possibilité de compromettre les efforts déployés dans l'atteinte de l'objectif 13, en raison des besoins énergétiques élevés des SIA, surtout si des énergies non carboneutres sont utilisées¹⁰⁹.

Afin de compléter le survol des utilisations possibles qu'offre l'IA dans la lutte contre les changements climatiques, il convient de s'attarder à l'[étude à grande échelle](#) menée par David Rolnick et ses collègues. Cette dernière a permis d'identifier 37 cas d'utilisation dans 13 domaines où l'IA peut avoir un impact important dans la lutte contre les changements climatiques (voir l'annexe). Cette étude appuie les autres travaux démontrant le potentiel de l'IA pour répondre aux défis liés aux enjeux environnementaux et aux changements climatiques.

Toutefois, le déploiement de l'IA comme solution dans la lutte contre les changements climatiques présente des risques. Au-delà de l'impact environnemental direct dont il a été question précédemment, chaque déploiement de

¹⁰⁵ Global Partnership on Artificial Intelligence (GPAI)(2022).

¹⁰⁶ Vinuesa et autres (2020).

¹⁰⁷ Cowls et autres (2023); Kwock (2019); Vinuesa et autres (2020).

¹⁰⁸ Vinuesa et autres (2020).

¹⁰⁹ Les auteurs notent que bien que leur analyse s'appuie sur différentes sources arrivant aux mêmes conclusions, il faut prendre en considération le risque possible de biais dans ce type de recherche. Les intérêts personnels de la communauté scientifique et de l'industrie à publier des résultats positifs, principalement en matière d'objectifs de développement durable, sont à prendre en considération. Aussi, il faut reconnaître que les aspects négatifs de l'IA pourraient nécessiter des recherches à plus long terme et qu'il existe actuellement peu de méthodes d'évaluation pour le faire.



SIA engendre des risques qui peuvent avoir un impact sur les individus, les organisations, la société, tous domaines confondus. Ces risques peuvent être classifiés de diverses façons, mais sont généralement liés à la performance, la sécurité, le contrôle, l'éthique, l'économie et la société¹¹⁰. Le cas de l'utilisation de l'IA dans la lutte contre les changements climatiques n'y fait pas exception; les risques existants sont seulement appelés à prendre une nouvelle forme¹¹¹.

On observe aussi des effets négatifs lorsque le déploiement de l'IA ne présente pas de valeur ajoutée, lorsqu'elle est utilisée incorrectement et mène à des résultats incorrects ou biaisés, ou lorsqu'elle est utilisée pour faire de l'écoblanchiment afin de détourner l'attention du besoin de changement fondamental du modèle d'affaires¹¹². Les risques environnementaux indirects de l'IA peuvent également comprendre son utilisation pour soutenir des modèles d'affaires non durables ou le commerce électronique et, ainsi, contribuer à l'augmentation de la demande en ressources, ce qui a un impact important sur l'environnement et la biodiversité¹¹³. Ces types d'impacts sont explicités dans la section suivante.

1.2.2. *Impacts indirects négatifs*

L'utilisation de SIA peut engendrer des impacts négatifs si ces derniers soutiennent des modèles d'affaires non durables. Ils peuvent, par exemple, exacerber les impacts environnementaux des secteurs manufacturiers et de l'exploitation minière. Bien que l'IA puisse être utilisée pour soutenir les efforts en matière de durabilité et contribuer à accroître les gains d'efficacité, elle peut aussi contribuer à augmenter les émissions nettes de GES plutôt qu'à les réduire. Les systèmes de recommandation propulsés par l'IA dans le domaine du commerce électronique représentent un autre exemple qui pourrait accroître une consommation atteignant des niveaux non viables¹¹⁴.

Dans un autre ordre d'idées, tenant compte du fait que l'IA accélère actuellement le développement technologique, il est possible que ces systèmes engendrent des phénomènes de sentier de dépendance (*path dependencies*). Il s'agit de la tendance des institutions ou des technologies à s'engager et se développer d'une certaine manière, en raison de leurs propriétés structurelles ou de leurs croyances et valeurs. Ce phénomène affecte l'atténuation des changements climatiques. Parallèlement, la littérature soulève aussi un effet de *lock-in* applicable aux situations dites d'éco-innovation. Le concept de *lock-in* fait référence à un scénario dans lequel une technologie s'impose en premier au sein d'un marché et qu'elle empêche les concurrents d'y accéder, alors qu'elle présente de potentielles inefficacités. Une éco-innovation, bien qu'il en existe diverses définitions, est une innovation qui se traduit par une réduction de l'impact sur l'environnement, qu'il soit intentionnel ou non¹¹⁵.

On qualifie l'IA d'éco-innovation lorsqu'on l'utilise dans un contexte de lutte contre les changements climatiques, de telle sorte que les pertes d'efficacité qu'elle engendre sur le marché ne sont pas prises en compte comme dans le cas d'innovations courantes. Dans le cas des éco-innovations, les inefficacités sont importantes d'un point de vue environnemental, social et institutionnel, et donc plus difficiles à calculer. Néanmoins, elles doivent être prises en compte dans l'établissement de modèles de croissance durables. [Cette même étude](#) avance qu'il est important de comprendre les processus sociaux, institutionnels et technologiques qui renforcent ces phénomènes de sentier de dépendance dans l'élaboration de politiques environnementales durables, considérant l'historique de bon nombre de technologies dominantes ayant eu des effets néfastes sur l'environnement, limitant ainsi le potentiel d'une croissance économique durable.

Le phénomène de *lock-in*, sous l'angle de l'empreinte carbone de l'IA, est aussi soulevé dans un [article](#) publié par Scott Robbins et Aimee van Wynsberghe dans *Sustainability*. Leur argumentaire repose sur le fait que l'IA s'insère

¹¹⁰ FEM (2018).

¹¹¹ *Ibid.*

¹¹² GPAI (2022).

¹¹³ *Ibid.*

¹¹⁴ OCDE, 2022a

¹¹⁵ OCDE (2009).



aujourd’hui dans plusieurs infrastructures fondamentales de notre société (transport, éducation, énergie, etc.), pour tantôt les « protéger », tantôt les améliorer. Selon les auteurs, l’IA ne s’insère plus seulement dans les infrastructures, mais devient une infrastructure en soi sur laquelle vont s’appuyer les services d’aujourd’hui et de demain. Puisque les SIA génèrent des impacts négatifs notables sur l’environnement, les auteurs soutiennent qu’on se dirige vers une situation de *lock-in carbone* qui viendrait restreindre notre habileté à réduire notre empreinte carbone sur le long terme. La course à l’adoption de l’IA dans chaque sphère de notre société aurait pour conséquence de créer des dépendances et des interdépendances, desquelles il serait ensuite difficile de se libérer, d’où l’importance de bien calculer les différents impacts avant de s’avancer dans ce *lock-in*.

Finalement, le phénomène d’effet rebond est abordé à plusieurs reprises dans la littérature portant sur les impacts environnementaux de l’IA. L’effet rebond, comme l’indique le paradoxe de Jevon, se produit lorsqu’un gain d’efficacité réalisé en réponse à un progrès technologique est compensé par une accélération de la consommation des ressources¹¹⁶. Bien qu’on retrouve une vaste quantité d’articles relatant l’impact positif que peut avoir l’IA sur les gains de productivité et d’efficacité dans les différents secteurs et chaînes d’approvisionnement, ce constat n’est pas unanime au sein de la communauté scientifique.

Peter Dauvergne¹¹⁷ avance que les gains réels en matière de durabilité environnementale ne sont pas aussi importants que le laisseraient croire les sociétés transnationales. En effet, les gains établis dans les chaînes d’approvisionnement seraient déplacés vers une production et une consommation accrue, faisant ainsi profiter économiquement ces dernières¹¹⁸. Un autre exemple d’effet rebond structurel à plus large impact est celui des véhicules autonomes propulsés par l’IA. Ces derniers amélioreraient certes l’efficacité énergétique, mais pourraient en même temps mener vers une augmentation des déplacements individuels en voiture, ce qui ferait augmenter directement la consommation d’énergie et les émissions si les véhicules ne sont pas partagés ou électrifiés¹¹⁹.

Dans le même ordre d’idées, la forte demande énergétique des SIA, et des technologies de l’information et des communications (TIC) plus largement, a mené certains acteurs de la communauté scientifique à s’interroger sur la contradiction qu’ils engendrent face à la demande croissante d’une sobriété numérique¹²⁰. La sobriété énergétique est « un principe qui vise à rendre l’utilisation du numérique compatible avec les limites planétaires, tout en conservant un accès équitable aux bénéfices qu’il génère (accès à l’information, aux ressources de soins de santé, d’éducation, etc.), et ce, pour nos contemporains ainsi que pour les générations futures »¹²¹.

Comme l’indique le CEST, les TIC sont déjà responsables d’émettre des GES et de consommer beaucoup de ressources naturelles, tout comme les SIA¹²² pour lesquels la demande ne cesse de croître. Cela a mené certains chercheurs comme Romain Couillet et ses collègues à tenter de comprendre l’impact d’un tel phénomène sur le principe de sobriété énergétique.

Tandis que les possibilités de développement de SIA ne cessent de se multiplier, les impacts environnementaux demeurent bien réels. Outre l’énorme demande énergétique nécessaire à l’entraînement de SIA¹²³, l’un des principaux moteurs d’incompatibilité avec le principe de sobriété numérique est le coût environnemental de production de la multiplication de dispositifs pour l’utilisateur final qui résultent de cette génération de nouveaux usages et besoins liés à l’IA¹²⁴. Aux émissions de GES s’ajoutent la contamination des sols et de l’eau, et l’épuisement des combustibles

¹¹⁶ Giampietro et Mayumi (2018).

¹¹⁷ Dauvergne (2020).

¹¹⁸ Dauvergne (2020); Kaack et autres (2022).

¹¹⁹ Kaack et autres (2022).

¹²⁰ Couillet et autres (2022).

¹²¹ CEST (2023).

¹²² Luccioni et autres (2022); OCDE (2022).

¹²³ Luccioni et autres (2022).

¹²⁴ Couillet et autres (2022)



fossiles et des minéraux comme les terres rares, qui sont des conséquences importantes du déploiement de SIA et rendent difficile l'atteinte d'une sobriété numérique¹²⁵.

En réponse à ces enjeux, certaines avenues sont discutées pour tenter d'atteindre une sobriété numérique. On parle notamment de priorisation des usages. Cela comprend toutes les activités que permet le numérique, de la restriction à certains contenus de diffusion vidéo continue à la limitation de l'énergie allouée aux cryptomonnaies¹²⁶ en passant par la promotion d'outils et le développement de laboratoires dits « *low-tech* ». La *low-tech* consiste en la fabrication de technologies en faisant une utilisation minimaliste, simple et durable des ressources (recyclage ou réutilisation de matériaux, utilisation d'énergies renouvelables, etc.) avec un coût limité et une conception adaptée pour une résilience locale¹²⁷. Pour parvenir à la sobriété numérique, l'action doit être individuelle, mais elle doit également émaner d'une volonté nationale et internationale.

2. Rôle des normes internationales

Les normes internationales présentent un potentiel notable en matière de standardisation des SIA. En 2017, l'Organisation internationale de normalisation (ISO) s'est associée à la Commission électrotechnique internationale (IEC) pour le développement de normes en IA. Depuis, ces normes exercent une influence en matière de recommandations et de régulation de l'IA. Avec les différents projets de législation et de réglementation des pays qui ne tarderont pas à être adoptés, les normes « représentent globalement une évolution vers l'opérationnalisation de la réglementation de l'IA, en donnant à ceux qui déploient et réglementent les systèmes d'IA les processus et les outils techniques nécessaires pour assurer la conformité réglementaire »¹²⁸.

On recense déjà plusieurs normes applicables aux SIA ou à la gestion environnementale du cycle de vie d'un produit :

- ISO 14040:2006 – Management environnemental – Analyse du cycle de vie – Principes et cadre;
- ISO 14067:2018 – Empreinte carbone des produits – Exigences et lignes directrices pour la quantification;
- ISO/IEC JTC 1/SC 42 – Applications d'intelligence artificielle (comité de travail);
- ISO/IEC 22989:2022 – Technologies de l'information – Intelligence artificielle – Concepts et terminologie relatifs à l'intelligence artificielle;
- ISO/IEC 23894:2023 – Technologies de l'information – Intelligence artificielle – Recommandations relatives au management du risque.

Un sous-comité (ISO/IEC JTC 1/SC 42) a été mis sur pied afin de s'attaquer à la question de l'IA. Toutefois, on dénote d'ores et déjà certains enjeux dans le développement de normes pour l'encadrer. En effet, les législateurs peinent à s'entendre sur les définitions des principes phares de l'IA responsable que sont l'équité, la robustesse, la transparence et le *droit de recours*. Cela engendre certaines incertitudes sur *la manière de mettre en pratique et de valider ces principes phares*¹²⁹.

De plus, les organisations chargées d'élaborer ces normes proviennent majoritairement des industries. Leur expérience est donc davantage liée aux enjeux et aux questions d'ordre technique plutôt qu'éthique et social. Par conséquent, et comme le soulève l'[article](#) d'Hadrien Pouget publié par l'*Ada Lovelace Institute*, on doit se demander si les organismes de normalisation détiennent la capacité institutionnelle pour créer les normes relatives à une technologie comme l'IA, vis-à-vis de laquelle les jugements de valeur et les enjeux sociaux dépassent les seules questions techniques. Bien qu'il ne fasse aucun doute que les normes puissent contribuer à la standardisation des

¹²⁵ *Ibid.*

¹²⁶ CEST (2023).

¹²⁷ Couillet et autres (2022).

¹²⁸ Pouget (2023).

¹²⁹ Pouget (2023).



SIA, il faut se demander ce qui devrait être standardisé ou non, mais aussi *qui a la légitimité de le faire et quel devrait être le rôle des normes internationales dans la gouvernance de l'IA*¹³⁰.

3. Survol d'initiatives d'encadrement hors Québec

Les gouvernements sont de plus en plus sensibilisés aux enjeux liés à l'IA et à son interaction avec les changements climatiques. Par exemple, la législation sur l'IA¹³¹ exige que tout SIA à haut risque évalue et signale l'impact négatif et raisonnablement prévisible de son utilisation sur l'environnement. La loi appelle également à la création d'indicateurs clés de performance pour suivre la consommation d'énergie des systèmes et promouvoir l'utilisation de technologies d'IA plus efficaces. La création de tels indicateurs permettra aussi de mesurer leur impact sur les objectifs de développement durable de l'ONU.

Par ailleurs, les États-Unis et l'Union européenne ont récemment annoncé¹³² qu'ils collaboreront pour utiliser l'IA afin de relever les grands défis mondiaux dans cinq domaines clés : les prévisions météorologiques et climatiques extrêmes; la gestion des interventions d'urgence; l'amélioration de la santé et de la médecine; l'optimisation du réseau électrique; l'optimisation de l'agriculture. Quatre d'entre eux sont directement liés aux changements climatiques, ce qui envoie un message clair sur la façon dont le secteur public envisage d'utiliser cette technologie dans les prochaines décennies.

Finalement, en mars 2023, le gouvernement britannique a confirmé un investissement de 900 millions de livres sterling en recherche sur l'IA, qui sera entre autres utilisée pour mieux comprendre les changements climatiques¹³³.

¹³⁰ *Ibid.*

¹³¹ Commission européenne (2021).

¹³² The White House (2023).

¹³³ Wald et Zhang (2023).



Annexe de la section Impact environnemental et intelligence artificielle

	Causal inference	Computer vision	Interpretable models	NLP	RL & Control	Time-series analysis	Transfer learning	Uncertainty quantification	Unsupervised learning
1 Electricity systems									
Enabling low-carbon electricity	•	•		•	•			•	•
Reducing current-system impacts	•				•			•	•
Ensuring global impact	•						•		•
2 Transportation									
Reducing transport activity	•			•	•			•	•
Improving vehicle efficiency	•			•					
Alternative fuels & electrification				•					•
Modal shift	•	•		•	•		•		
3 Buildings and cities									
Optimizing buildings	•		•	•	•				
Urban planning		•	•	•	•				
The future of cities			•	•	•		•	•	•
4 Industry									
Optimizing supply chains	•		•	•					
Improving materials									•
Production & energy	•	•	•	•					
5 Farms & forests									
Remote sensing of emissions	•								
Precision agriculture	•		•	•					
Monitoring peatlands	•								
Managing forests	•		•	•					
6 Carbon dioxide removal									
Direct air capture									•
Sequestering CO ₂	•						•	•	
7 Climate prediction									
Uniting data, ML & climate science	•	•		•				•	
Forecasting extreme events	•	•		•				•	
8 Societal impacts									
Ecology	•						•		
Infrastructure			•	•	•			•	
Social systems	•		•	•	•				
Crisis	•		•	•	•				•
9 Solar geoengineering									
Understanding & improving aerosols								•	
Engineering a planetary control system				•	•			•	
Modeling impacts				•	•			•	
10 Individual action									
Understanding personal footprint	•		•	•	•	•			
Facilitating behavior change			•	•	•				•
11 Collective decisions									
Modeling social interactions		•	•	•	•				
Informing policy	•	•	•	•	•			•	
Designing markets			•	•	•			•	
12 Education									
13 Finance								•	

Domaines d'impact climatique positif potentiel et formes de technologie d'IA pertinentes pour chacun (Rolinck et autres [2019]).



Bibliographie de la section Impact environnemental et intelligence artificielle

Cecere, G., Corrocher, N., Gossart, C., et Ozman, M. (2014). *Lock-in and path dependence: An evolutionary approach to eco-innovations*. Journal of Evolutionary Economics, 24(5), 1037-1065. <https://doi.org/10.1007/s00191-014-0381-5>.

CEST. (2023). *La sobriété numérique et la question de la priorisation des usages collectifs*. Commission de l'éthique en science et en technologie.

<https://www.ethique.gouv.qc.ca/fr/actualites/ethique-hebdo/la-sobriete-numerique-et-la-question-de-la-priorisation-des-usages-collectifs/>.

Commission européenne. (2021). *Proposal for a Regulation on Artificial Intelligence [COM/2021/206 final]*. Brussels, Belgium: Publications Office of the European Union.

Couillet, R., Trystram, D., et Menissier, T. (2022). *The Submerged Part of the AI-Ceberg [Perspectives]*. IEEE Signal Processing Magazine, 39(5), 10-17.

<https://doi.org/10.1109/MSP.2022.3182938>.

Cowls, J., Tsamados, A., Taddeo, M., et Floridi, L. (2023). *The AI gambit: Leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations*. AI & SOCIETY, 38(1), 283-307. <https://doi.org/10.1007/s00146-021-01294-x>.

Dauvergne, P. (2020). *Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs*. Review of International Political Economy, 29(3), 696-718. <https://doi.org/10.1080/09692290.2020.1814381>.

Dennis, C., Leon Kirnberger, J., et Shankar, V. (30 mai 2023). *We need to use AI to fight climate change—OECD.AI*. The AI Wonk. <https://oecd.ai/en/wonk/fight-climate-change>.

Duberry, J. (2021). *Intelligence artificielle, données volumineuses et conservation de la biodiversité*. Vertigo – la revue électronique en sciences de l'environnement, Volume 21 numéro 2. <https://journals.openedition.org/vertigo/32470>.

Giampietro, M., et Mayumi, K. (2018). *Unravelling the Complexity of the Jevons Paradox : The Link Between Innovation, Efficiency, and Sustainability*. Frontiers in Energy Research, 6, 26. <https://doi.org/10.3389/fenrg.2018.00026>.

Global Partnership on Artificial Intelligence (GPAI). (2022). *Biodiversity and AI: Opportunities and Recommendations for Action*. Global Partnership on AI. <https://gpai.ai/projects/responsible-ai/environment/biodiversity-and-AI-opportunities-recommendations-for-action.pdf>.

Ham, Y.-G., Kim, J.-H., et Luo, J.-J. (2019). *Deep learning for multi-year ENSO forecasts*. Nature, 573(7775), Article 7775. <https://doi.org/10.1038/s41586-019-1559-7>.

Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., et Yang, H. (2019). *Machine learning and artificial intelligence to aid climate change research and preparedness*. Environmental Research Letters, 14(12), 124007. <https://doi.org/10.1088/1748-9326/ab4e55>.

IEA. (2022). *Data Centres and Data Transmission Networks*. IEA. <https://www.iea.org/reports/digitalisation-and-energy>.



Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., et Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 12(6), Article 6. <https://doi.org/10.1038/s41558-022-01377-7>.

Kwok, R. (2019). AI empowers conservation biology. *Nature*, 567(7746), 133-134. <https://doi.org/10.1038/d41586-019-00746-1>.

Lagueux-Beloin, A. (2019). Une tonne équivalent CO₂, c'est gros comment et ça représente quoi? <https://unpointinq.ca/comprendre/tonne-equivalent-co2/>.

Luccioni, A. S., Viguier, S., et Ligozat, A.-L. (2022). Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model(arXiv:2211.02001). arXiv. <http://arxiv.org/abs/2211.02001>.

Ludvigsen, K. G. A. (2022). How to estimate and reduce the carbon footprint of machine learning models. Medium. <https://towardsdatascience.com/how-to-estimate-and-reduce-the-carbon-footprint-of-machine-learning-models-49f24510880>.

Narciso, D. A. C., et Martins, F. G. (2020). Application of machine learning tools for energy efficiency in industry: A review. *Energy Reports*, 6, 1181-1199. <https://doi.org/10.1016/j.egyr.2020.04.035>.

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., et Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115(25), E5716-E5725. <https://doi.org/10.1073/pnas.1719367115>.

OECD. (2009). Sustainable Manufacturing and Eco-innovation: Towards a Green Economy. <https://www.oecd.org/env/consumption-innovation/42957785.pdf>.

OCDE. (2022a). Measuring the environmental impacts of artificial intelligence compute and applications : The AI footprint. OCDE. <https://doi.org/10.1787/7babf571-en>.

OCDE. (2022b). OECD Framework for the Classification of AI systems. <https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm>.

Pouget, H. (2023). What will the role of standards be in AI governance? Ada Lovelace Institute. <https://www.adalovelaceinstitute.org/blog/role-of-standards-in-ai-governance/>.

Robbins, S., et van Wynsberghe, A. (2022). Our New Artificial Intelligence Infrastructure: Becoming Locked into an Unsustainable Future. *Sustainability*, 14(8), Article 8. <https://doi.org/10.3390/su14084829>.

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C., Ng, A. Y., Hassabis, D., Platt, J. C., ... Bengio, Y. (2019). Tackling Climate Change with Machine Learning (arXiv:1906.05433). arXiv. <https://doi.org/10.48550/arXiv.1906.05433>.

Silvestro, D., Goria, S., Sterner, T., et Antonelli, A. (2022). Improving biodiversity protection through artificial intelligence. *Nature Sustainability*, 5(5), Article 5. <https://doi.org/10.1038/s41893-022-00851-6>.



The Shifters Montréal. (2020). *L'impact environnemental du numérique au Québec et au Canada* (deuxième rapport du projet DiagnosTIC).

The White House. (2023). *Statement by National Security Advisor Jake Sullivan on the New U.S.-EU Artificial Intelligence Collaboration* (communiqué de presse).

<https://www.whitehouse.gov/briefing-room/statements-releases/2023/01/27/statement-by-national-security-advisor-jake-sullivan-on-the-new-u-s-eu-artificial-intelligence-collaboration/>.

Thilakarathna, P. S. M., Seo, S., Baduge, K. S. K., Lee, H., Mendis, P., et Foliente, G. (2020). *Embodied carbon analysis and benchmarking emissions of high and ultra-high strength concrete using machine learning algorithms*. Journal of Cleaner Production, 262, 121281.

<https://doi.org/10.1016/j.jclepro.2020.121281>.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., et Fuso Nerini, F. (2020). *The role of artificial intelligence in achieving the Sustainable Development Goals*. Nature Communications, 11(1), Article 1.

<https://doi.org/10.1038/s41467-019-14108-y>.

Wald, R., et Zhang, D. (2023). *Conservative British Government Is Seizing the Day on AI While the U.S. Dawdles*. Stanford HAI. <https://hai.stanford.edu/news/conservative-british-government-seizing-day-ai-while-us-dawdles>.

World Economic Forum. (2018). *Harnessing Artificial Intelligence for the Earth report 2018*. (Fourth Industrial Revolution for the Earth Series).

https://www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf.



III. Intelligence artificielle, arts et culture

L'utilisation croissante de l'IA au sein de la société transforme la culture même de cette dernière, en plus de remettre en question notre rapport aux arts et à la culture. Plus récemment, l'amélioration fulgurante de l'IA générative est venue accentuer ce phénomène. Soulignons d'abord que les SIA affectent les arts et la culture de diverses façons. En effet, ils transforment notamment la production, la circulation ainsi que la consommation culturelle¹³⁴. En plus des enjeux propres à la diffusion et à la création artistique par l'IA, il est aussi question d'enjeux touchant plus généralement l'appréciation et la compréhension de l'art¹³⁵. Nous avons donc regroupé les différents enjeux relatifs aux impacts de l'IA sur les arts et la culture en trois catégories :

- 1) Diffusion et consommation des contenus culturels;
- 2) Création des contenus culturels;
- 3) Perception et compréhension des contenus culturels.

À noter que nous nous basons sur une conception large de la culture telle que la définit par exemple l'UNESCO, soit « l'ensemble des traits distinctifs, spirituels et matériels, intellectuels et affectifs, qui caractérisent une société ou un groupe social. Elle englobe, outre les arts et les lettres, les modes de vie, les droits fondamentaux de l'être humain, les systèmes de valeurs, les traditions et les croyances »¹³⁶. Cependant, nous effectuons tout de même une distinction entre « culture » et « arts » afin de mettre l'accent sur des éléments spécifiques de cette définition. Le terme « culture » (ou les « cultures ») fera donc davantage référence aux « traits distinctifs qui caractérisent une société ou un groupe », par exemple la langue, ou au groupe qui partage ces traits. Les « arts » (ou « l'art ») correspondent plutôt à une forme centrale de production et d'expression de cette culture. Malgré cette compréhension large du concept de culture, nous nous concentrerons, sans toutefois nous y restreindre, aux enjeux se rapportant aux « contenus culturels », c'est-à-dire à des contenus renvoyant « au sens symbolique, à la dimension artistique et aux valeurs culturelles qui ont pour origine ou expriment des identités culturelles »¹³⁷.

1. Diffusion et consommation des contenus culturels

Cette première catégorie permet de prendre en compte l'impact des algorithmes des grandes plateformes numériques comme Spotify, Netflix ou Prime sur l'offre et la diversité des contenus culturels. Pour des raisons essentiellement commerciales, ces géants du numérique ont rapidement intégré l'IA à leurs plateformes¹³⁸. Dans un contexte où la diffusion de contenus culturels se fait de plus en plus au moyen de ces plateformes, les suggestions et recommandations influencent fortement le contenu que nous consommons. Le rôle joué par les algorithmes affecte notamment notre capacité à découvrir des contenus culturels, ce qui renvoie au concept de découvrabilité. Plus spécifiquement, on peut la définir de la manière suivante :

*La découvrabilité d'un contenu dans l'environnement numérique se réfère à sa disponibilité en ligne et à sa capacité à être repéré parmi un vaste ensemble d'autres contenus, notamment par une personne qui n'en faisait pas précisément la recherche.*¹³⁹

Au-delà de la découvrabilité et de l'utilisation d'algorithmes pour faire des suggestions de contenu aux utilisateurs, les grandes plateformes utilisent également ces derniers afin de déterminer, à partir des données des utilisateurs, les prochains contenus à produire.

¹³⁴ Jin (2021a).

¹³⁵ Cetinic et She (2022).

¹³⁶ UNESCO (1982).

¹³⁷ UNESCO (2005).

¹³⁸ Jin (2021a).

¹³⁹ Ministère de la Culture et des Communications du Québec et ministère de la Culture de France (2020).



Découvrabilité

Les grandes entreprises de diffusion sont guidées par une logique de profitabilité à laquelle s'ajoute une mondialisation culturelle toujours croissante. Les algorithmes de recommandation qu'elles utilisent créent un accès inéquitable et non diversifié au marché mondial pour les œuvres d'artistes nationaux et locaux. Il est en ce sens possible de parler d'une forme de monopole de la découvrabilité et d'une homogénéisation de l'offre culturelle¹⁴⁰. Cette « nouvelle » offre est par ailleurs principalement dominée par la culture américaine¹⁴¹.

Puisque ce sont les contenus anglophones qui prédominent massivement sur ces plateformes, la protection et la préservation des langues sont ici en jeu. Dans le contexte québécois, c'est entre autres la langue française qui est menacée. À ce sujet, une étude relève même que la découvrabilité des contenus musicaux et audiovisuels franco-québécois sur les grandes plateformes, malgré leur présence, est pratiquement inexistante¹⁴². Dans une perspective plus globale, le [rapport de recherche](#) de Destiny Tchekhouali et de Christian Agboblé indique que la découvrabilité des contenus francophones en ligne est faible à l'échelle de la planète, en plus de montrer que les algorithmiques produisent des erreurs et des biais dans leurs recommandations pour les utilisateurs francophones en faveur de contenus musicaux « d'artistes-vénettes de la musique pop occidentale ».

Plusieurs initiatives du gouvernement du Québec concernant la découvrabilité ont d'ailleurs déjà été lancées. Mentionnons notamment les différents travaux réalisés dans le cadre de la Mission sur la découvrabilité en ligne des contenus culturels francophones, menée conjointement par les ministères de la culture québécois et français. Une stratégie commune a notamment été élaborée. Cela a conduit, entre autres, à la mise en place d'un programme de soutien financier pour des projets – incluant des projets de recherche – qui visent à favoriser la découvrabilité des contenus culturels francophones en ligne. Plus récemment, le ministère de la Culture et des Communications du Québec (MCC) a formé le Groupe d'experts pour un meilleur rayonnement de la culture québécoise et de la langue française sur les plateformes numériques. Ce groupe aura pour objectif d'analyser les outils et les approches afin « d'améliorer le positionnement du contenu culturel québécois francophone sur les grandes plateformes numériques »¹⁴³.

À noter que le gouvernement fédéral, avec la *Loi sur la diffusion continue en ligne* adoptée en avril 2023, réglemente les enjeux de découvrabilité sur les grandes plateformes. En effet, cette loi « vise à mettre à jour la *Loi sur la radiodiffusion*, afin d'obliger les plateformes numériques telles que Netflix, YouTube et TikTok à promouvoir le contenu canadien et à y contribuer »¹⁴⁴. Cette loi placera également la réglementation des plateformes sous l'autorité du Conseil de la radiodiffusion et des télécommunications canadiennes (CRTC), qui aura comme objectif de soutenir et valoriser le contenu canadien et autochtone¹⁴⁵.

Alors que la question des contenus autochtones est présente dans cette initiative fédérale, elle semble l'être moins dans les initiatives entreprises par l'État québécois. En effet, les mesures à l'échelle québécoise sont essentiellement centrées sur la découvrabilité des contenus francophones d'ici, et peu sur les contenus culturels produits et consommés par les onze nations autochtones se trouvant sur le territoire québécois, qui subissent pourtant tout autant les effets de l'uniformisation de l'offre culturelle. L'accès aux œuvres et expressions culturelles en langues autochtones dépend de « leur disponibilité et leur visibilité dans la sphère publique, ainsi que leur découvrabilité dans l'environnement numérique »¹⁴⁶.

Plusieurs solutions et recommandations afin de favoriser la découvrabilité ont déjà été émises dans les multiples travaux menés jusqu'à maintenant. Parmi celles-ci, on retrouve par exemple l'idée de réguler les grandes plateformes

¹⁴⁰ Guèvremont (2021).

¹⁴¹ Jin (2021); Tchekhouali et Agboblé (2020a et 2020b).

¹⁴² Rioux et autres (2021a).

¹⁴³ Cabinet du ministre de la Culture et des Communications et du ministre responsable de la Jeunesse (2023).

¹⁴⁴ La Presse canadienne (2023).

¹⁴⁵ Radio-Canada (2023).

¹⁴⁶ Mariage et Guèvremont (2022).



numériques¹⁴⁷ et d'imposer des quotas, ou encore d'envisager la création de plateformes numériques alternatives¹⁴⁸. L'idée de taxer les plateformes numériques étrangères pour réinvestir dans la culture locale, puis de soutenir le rôle des médias dans la promotion de la culture¹⁴⁹ est également avancée.

Détermination des contenus à produire

Les algorithmes des grandes plateformes numériques ne font toutefois pas des recommandations uniquement aux utilisateurs. Ils en font également de plus en plus aux producteurs de contenu, c'est-à-dire aux plateformes elles-mêmes. Autrement dit, les algorithmes, à partir des préférences et des habitudes des utilisateurs, sont aussi utilisés pour déterminer les prochains contenus culturels qui seront produits. Netflix, avec son rôle de diffuseur et de producteur de contenu, a été la première grande plateforme à adopter cette stratégie, notamment avec la série *House of Cards*. Depuis, cette stratégie a été reprise par les autres grands joueurs de l'industrie¹⁵⁰.

Tout comme pour les enjeux liés à la découvervabilité, cette pratique homogénéise l'offre culturelle et menace ainsi la diversité des expressions culturelles¹⁵¹. En effet, cette utilisation des algorithmes conduit à la production de contenus imitant ce qui a déjà été fait précédemment dans le but de plaire aux utilisateurs, ce qui nuit à la créativité, à la qualité et à la diversité des contenus culturels¹⁵².

2. Crédit des contenus culturels

L'IA n'affecte cependant pas seulement la manière dont les contenus culturels sont diffusés et consommés, mais également la manière dont ils sont créés. Les modèles d'IA générative se sont grandement et rapidement perfectionnés dans les dernières années, voire les derniers mois, notamment grâce à l'amélioration des capacités de traitement du langage. En effet, plusieurs modèles sont maintenant en mesure de produire presque instantanément du contenu « nouveau » et complexe, que ce soit sous forme de texte (ChatGPT, PaLM), d'illustration (Dall-E, Midjourney) ou même de son et de musique (VALL-E, ressemble.ai).

Propriété intellectuelle et droit d'auteur

Depuis quelques années déjà, les créations générées par des SIA, ou à l'aide de ceux-ci, soulèvent des interrogations en matière d'application des règles de propriété intellectuelle et de droit d'auteur. En effet, le contenu produit par ces modèles est tiré de bases de données contenant les créations et les œuvres de milliers de personnes n'ayant pas nécessairement consenti à ce que soit utilisé le fruit de leur travail par ces modèles. C'est entre autres la phase d'entraînement qui est au cœur de ces enjeux. En effet, c'est durant cette phase qu'est incorporée une quantité massive d'images, de travaux écrits et de sons¹⁵³. Dernièrement, plusieurs auteurs ont intenté des poursuites contre les grandes entreprises derrière ces modèles à propos de l'intégration, sans leur consentement, de leurs œuvres et textes¹⁵⁴.

Au-delà de la phase d'entraînement et de l'incorporation des données, la création de contenu culturel par l'IA met aussi en lumière des enjeux concernant la propriété intellectuelle de ce contenu. L'idée d'accorder la propriété du contenu généré par l'IA à l'IA elle-même est parfois abordée¹⁵⁵. Toutefois, cela ne serait par exemple pas possible actuellement selon la loi américaine, puisque le droit d'auteur s'applique uniquement aux créations produites par des êtres

¹⁴⁷ Rioux et autres (2021b).

¹⁴⁸ Guèvremont et autres (2019); Tchekhouali et Agbobji (2020a).

¹⁴⁹ Guèvremont et autres (2019).

¹⁵⁰ Jin (2021).

¹⁵¹ Guèvremont (2021).

¹⁵² Jin (2021).

¹⁵³ Torrance et Tomlinson (2023).

¹⁵⁴ Small (2023); Creamer (2023).

¹⁵⁵ Flick et Worrall (2022); Murray (2022); Jin (2021).



humains¹⁵⁶. Il faudrait en ce sens que l'IA soit considérée comme une personne morale, ce qui s'avérerait fort complexe¹⁵⁷. De manière semblable, la jurisprudence canadienne indique que le droit d'auteur ne s'applique qu'aux êtres humains¹⁵⁸.

Le gouvernement fédéral a d'ailleurs lancé, en 2021, une consultation sur l'encadrement juridique de l'IA¹⁵⁹. Cette dernière se penche à la fois sur la question du droit d'utilisation des données pour l'entraînement des modèles ainsi que sur le statut du droit d'auteur des contenus générés par l'IA, dans le cadre d'une possible révision de la *Loi sur le droit d'auteur* en l'adaptant aux défis que pose l'IA.

Cependant, bien que des recommandations à propos de l'utilisation des données dans les phases d'entraînement aient découlé de ce rapport, aucune recommandation d'amendements à propos de la propriété des contenus générés par l'IA n'a été proposée¹⁶⁰. Similairement, Hunter Parsons et Vivian Cheng, dans un [article](#) publié par le cabinet juridique BLG, affirment que la nouvelle loi canadienne adoptée dans le cadre du projet de loi C-27, la *Loi sur l'intelligence artificielle et les données* (LIAD), n'aborde pas la question du droit d'auteur en ce qui a trait à l'utilisation commerciale des œuvres artistiques créées à l'aide de l'IA générative.

Métiers de la culture

L'amélioration fulgurante et récente de l'IA générative est venue bousculer plus largement la place et le rôle de l'être humain dans l'art. Plus concrètement, il est possible d'envisager la disparition de certains métiers de la culture, ou du moins leur transformation. Pour l'écriture, certains emplois liés à la rédaction de contenus plus simples ou plus répétitifs pourraient ultimement être remplacés par les modèles d'IA générative. De manière similaire, les images et créations musicales produites par ces modèles pourraient convenir à certaines applications et possiblement remplacer le travail effectué par des travailleurs du milieu des arts visuels ou de la musique¹⁶¹.

Alors qu'il est souvent avancé que l'IA remplacera principalement les tâches plus répétitives, le déploiement des modèles de génération d'images menacerait dorénavant les emplois créatifs et spécialisés dans le domaine des arts visuels. Ces emplois incluent, bien sûr, les métiers plus classiques comme l'illustration et le design, mais également les métiers liés aux effets spéciaux et à l'animation dans l'industrie du cinéma et des jeux vidéo¹⁶². On peut également compter certains risques pour les emplois dans le domaine de la narration et du doublage. À cet égard, [l'Union des artistes a tout dernièrement fait part de ses inquiétudes](#) quant aux conséquences des nouveaux modèles d'IA sur l'avenir des artistes de la voix. Plus récemment, la *Writers Guild of America* est entrée en grève à Hollywood, notamment en raison des risques et des enjeux relatifs à l'IA générative. En effet, la guilde des auteurs, maintenant soutenue par celle des acteurs, soutient que les modèles d'IA générative menacent l'existence même du métier d'auteur. Elle souhaite également prévenir l'incorporation sans consentement du travail des auteurs dans la phase d'entraînement de ces modèles¹⁶³, renvoyant ainsi aux enjeux de droit d'auteur.

Malgré ces inquiétudes, les modèles d'IA générative peuvent également faciliter le travail de certains artistes et, ultimement, améliorer leurs créations. En effet, l'intégration de plusieurs formes d'art au sein de ces modèles pourrait être bénéfique pour l'exploration artistique et jouer un rôle important dans le processus créatif des artistes¹⁶⁴. De plus, beaucoup de ces modèles sont d'ailleurs conçus en impliquant les artistes dans le processus de développement,

¹⁵⁶ Murray (2022).

¹⁵⁷ Flick et Worrall (2022).

¹⁵⁸ Gouvernement du Canada (2021).

¹⁵⁹ *Ibid.*

¹⁶⁰ Macek (2023).

¹⁶¹ Flick et Worrall (2022).

¹⁶² Salkowitz (2022).

¹⁶³ Collier (2023).

¹⁶⁴ Cetinic et She (2022).



et ce, pour aider et soutenir plus globalement le travail des artistes¹⁶⁵. Le co-design peut donc être considéré comme un moyen de minimiser les différents risques liés au remplacement des métiers de la culture. Qui plus est, les humains continuent d'occuper un rôle central dans la production culturelle, car les modèles d'IA peinent à produire des contenus nouveaux et diversifiés¹⁶⁶.

Biais, stéréotypes et représentation de la diversité

La reproduction de biais et de stéréotypes est un des principaux enjeux relatifs au déploiement et à l'utilisation de l'IA. Les SIA peuvent effectivement contribuer à renforcer la marginalisation de certains groupes sociaux et à perpétuer certaines inégalités. Bien que ces enjeux aient déjà amplement été soulevés par le passé, ceux-ci prennent de formes nouvelles avec l'amélioration récente des modèles d'IA générative, notamment dans le domaine des arts et de la culture.

La reproduction de biais et de stéréotypes des modèles « text-to-image », par des IA telles que Dall-E, Stable Diffusion ou Midjourney, est par exemple déjà bien documentée. Dans un [article](#), Lukas Struppek et ses collègues démontrent qu'à cet égard, ces modèles reproduisent des « biais culturels », c'est-à-dire des caractéristiques et des stéréotypes ethniques et culturels. De manière semblable, un [article](#) de Federico Bianchi et ses collègues montrent que les images produites par ces modèles véhiculent des stéréotypes liés à « la race » et à la nationalité, mais également à d'autres éléments identitaires comme le genre, l'orientation sexuelle et même le statut socioéconomique.

À titre d'exemple, en demandant à Stable Diffusion une image d'une « personne attirante », le modèle montre seulement¹⁶⁷ des personnes « blanches ». Similairement, lorsqu'une image de « personne émotive » est demandée, des images de personnes d'apparence féminine sont générées¹⁶⁸. Une [étude](#) révèle aussi que la plupart des modèles de génération d'images produisent des biais de stéréotypes concernant le genre, la profession, la race et la religion, bien que les stéréotypes de genre soient légèrement plus présents que les autres. Ces modèles favorisent également par défaut la culture et les normes occidentales, plus spécifiquement celles des États-Unis¹⁶⁹.

La présence de ces biais dans le contenu généré par l'IA n'est toutefois pas limitée aux modèles de génération d'images. En effet, des modèles de génération de texte comme ChatGPT reproduisent également divers biais et stéréotypes culturels et démographiques¹⁷⁰. Cela est pertinent ici, car de tels modèles langagiers peuvent produire des textes considérés comme des formes de productions culturelles et artistiques. ChatGPT produirait également des biais de langage en favorisant l'anglais, et plus spécifiquement certaines formes d'anglais, par rapport à d'autres langues et formes d'écriture¹⁷¹.

Ce dernier élément montre non seulement que les modèles d'IA générative perpétuent des biais dans le contenu qu'ils génèrent, mais aussi que certains groupes culturels sont favorisés dans l'utilisation même de ces modèles. En effet, puisque le contenu à partir duquel se fait l'entraînement des grands modèles de langage comme ChatGPT se trouve sur Internet, il existe principalement en anglais et dans quelques autres langues dominantes. Cela fait en sorte que les modèles s'avèrent plus performants dans ces langues au détriment de langues et de dialectes minoritaires. Ce constat réduit en ce sens leur inclusion dans ces modèles. Pour assurer le partage équitable des avantages de l'IA

¹⁶⁵ Flick et Worrall (2022).

¹⁶⁶ Jin (2021).

¹⁶⁷ L'article montre 10 images choisies aléatoirement sur un total 100; les 10 correspondent à des personnes « blanches ».

¹⁶⁸ Bianchi et autres (2022).

¹⁶⁹ Struppek et autres (2022); Bianchi et autres (2022).

¹⁷⁰ Ferrara (2023).

¹⁷¹ Bjork (2023).



dans la société, il serait nécessaire de rendre ces modèles inclusifs et accessibles à l'ensemble des utilisateurs, et ce, sans discrimination de langue ou de culture¹⁷².

Différentes solutions techniques sont proposées pour pallier les différents biais et stéréotypes mis de l'avant par les modèles d'IA générative. À titre d'exemple, il existe des outils qui permettent de diminuer ces biais en intervenant sur les « *textual prompts* » des modèles de génération d'images¹⁷³. Bien que des outils et indicateurs puissent s'avérer utiles, le rôle du jugement humain dans le processus d'évaluation et de validation de ces biais demeure crucial, tout comme l'implication et l'inclusion des groupes marginalisés dans le développement et l'évaluation des modèles. Toutefois, la présence de certains biais dans les modèles d'IA générative, malgré les efforts et initiatives pour les empêcher reste inévitable. D'une part, les modèles sont entraînés sur des données produites par des individus; elles reflètent donc les biais qui existent dans la société. D'autre part, considérant la complexité du langage, certains biais sont plutôt liés aux capacités des modèles à traiter et à comprendre les demandes textuelles des utilisateurs. En ce sens, il est nécessaire de sensibiliser les utilisateurs de ces modèles à la présence de ces biais et aux effets potentiels de ceux-ci¹⁷⁴.

3. Perception et compréhension des contenus culturels

La troisième et dernière catégorie porte quant à elle sur l'influence de l'IA sur notre manière d'appréhender les arts et la culture, ce qui bouleverse plus largement notre rapport à ceux-ci. Il faut distinguer deux principaux éléments ici. Premièrement, l'utilisation croissante de l'IA pour produire des contenus artistiques et culturels transforme notre appréciation et notre perception des arts et de la culture. Deuxièmement, l'IA peut être utilisée pour analyser et interpréter les arts et la culture, notamment dans les domaines de l'histoire de l'art et de la muséologie. Bien que ces enjeux soient plus abstraits que les précédents, ils s'avèrent tout aussi cruciaux. Nous en explorerons les enjeux brièvement.

Perception et appréciation des nouvelles formes de contenus générés à l'aide de l'IA

L'intégration de l'IA aux processus de création et de production de divers contenus culturels transforme bien sûr les pratiques artistiques et culturelles, mais également la façon dont nous apprêhendons ces « nouveaux » contenus. Les chercheurs Joo-Wha Hong et Nathaniel Ming Curran ont mené une [étude](#) en 2019 dans le cadre de laquelle on a montré des images à 288 participants qui ignoraient que certaines d'entre elles étaient produites par un modèle d'IA. L'étude démontre que les œuvres visuelles créées par des humains étaient plus appréciées par les participants que celles créées par un modèle d'IA. Selon les chercheurs, ce résultat général est attribuable à une évaluation plus positive des aspects de « composition », de « degré d'expression » et de « valeur esthétique » pour les œuvres produites par des humains. Les auteurs concluent donc que l'IA ne semble pas encore avoir passé le test de Turing dans le domaine des arts visuels.

Cette étude remonte cependant à 2019. Depuis, la performance des modèles d'IA générative s'est grandement améliorée. Quelques récents exemples dans le monde des arts visuels semblent plutôt indiquer qu'il est de plus en plus difficile de distinguer les œuvres produites par des SIA de celles produites par des êtres humains. En mars dernier, Boris Eldagsen a remporté un prix aux Sony World Photography Awards en soumettant une photographie générée par DALL-E 2. L'artiste a cependant décliné le prix, mentionnant que son objectif était justement de déterminer si les compétitions d'art étaient préparées à l'arrivée des images générées par des SIA¹⁷⁵. De manière semblable, Jason M. Allen a gagné, en septembre dernier au Colorado, le prix de la meilleure œuvre dans la catégorie

¹⁷² Ferrara (2023).

¹⁷³ Zhang et autres (2023).

¹⁷⁴ Ferrara (2023); Bianchi et autres (2022).

¹⁷⁵ Parshall (2023).



« peinture » avec une œuvre produite par le modèle Midjourney. Toutefois, contrairement à Eldagsen, Allen a accepté le prix et s'est défendu de n'avoir enfreint aucune règle en ayant recours à un modèle d'IA¹⁷⁶.

Bien que l'accent ait été mis sur les arts visuels, la difficulté à distinguer les contenus culturels produits par les humains de ceux produits par des SIA concerne d'autres formes d'art et d'expression culturelle. Par exemple, il est difficile de déterminer avec certitude si du contenu textuel a été généré par un modèle de langage comme ChatGPT, puisque le but de ces modèles est justement d'imiter le langage humain¹⁷⁷. De plus, l'amélioration rapide des modèles d'IA générative risque d'étendre cette difficulté vers des domaines comme la musique ou l'animation. La capacité de ces modèles d'imiter les œuvres et le travail artistique réalisés par des humains altère notre manière même de concevoir l'art¹⁷⁸. Plus généralement, c'est le rôle et la place de l'artiste et de l'être humain dans l'art qui sont remis en question.

Utilisation de l'IA pour analyser les contenus culturels

L'IA est utilisée non seulement pour produire des contenus culturels, mais également en tant qu'outil pour analyser et comprendre ces derniers. Dans un [article](#) datant de 2022, les chercheurs Eva Cetinic et James She répertorient, au sein de la littérature, plusieurs utilisations possibles de l'IA pour « analyser l'art ». D'abord, ils démontrent que les algorithmes peuvent contribuer à construire des plateformes de recherche plus efficaces et des outils plus performants pour explorer les collections d'œuvres d'art dans un contexte où celles-ci sont de plus en plus numérisées. Ensuite, l'IA permet aussi d'étendre les possibilités en histoire de l'art en augmentant les capacités et les possibilités d'analyse des œuvres d'art numérisées dans les bibliothèques et collections numériques. Par exemple, l'IA peut être utilisée pour détecter des caractéristiques ou identifier des similarités entre différentes œuvres. Qui plus est, selon les auteurs, l'IA favoriserait le développement d'approches quantitatives en histoire de l'art, même si beaucoup de chercheuses et de chercheurs de ce champ sont méfiants vis-à-vis de la mouvance de la quantification des sciences humaines.

Toujours dans le même article, on indique que les algorithmes seraient également utilisés afin de mieux comprendre et même de prédire les jugements et les perceptions esthétiques des individus à propos de certaines œuvres en art visuel. Cependant, cela demeure assez difficile pour l'instant, notamment en raison de la complexité à quantifier et à prendre en compte des éléments comme l'originalité. Ici, on peut d'ailleurs faire un lien entre les développements de ces méthodes en histoire de l'art et l'utilisation des algorithmes par les grandes plateformes numériques dans le but de déterminer le contenu à produire pour satisfaire les utilisateurs (voir section 1.2). Si le développement de ces méthodes et approches prédictives en histoire de l'art peut contribuer à étendre les connaissances dans ce domaine, ces connaissances pourraient potentiellement être utilisées pour produire des contenus artistiques ciblés et, ultimement, diminuer l'originalité et la diversité de l'offre culturelle.

4. Survol des initiatives internationales

S'il existe plusieurs initiatives visant à encadrer les impacts de l'IA sur les arts et la culture, notamment au Québec et au Canada, il apparaît pertinent d'aborder brièvement les mesures et initiatives ayant d'ailleurs. Le Canada, avec le Royaume-Uni et l'Union européenne, fait partie des principaux États à avoir développé des politiques concernant la découverbarilité des contenus vidéo sur les grandes plateformes numériques. Malgré quelques différences, les stratégies et politiques de ces États insistent sur l'importance d'une régulation flexible pour s'adapter à l'évolution rapide de la technologie. En ce sens, ils proposent plutôt des principes généraux afin de laisser une certaine

¹⁷⁶ Roose (2022).

¹⁷⁷ Heikkilä (2023).

¹⁷⁸ Cetinic et She (2022).



autonomie décisionnelle aux autorités responsables de la réglementation qui, par ailleurs, collaborent souvent avec l'industrie¹⁷⁹.

Ensuite, pour ce qui est des contenus artistiques produits par l'IA, on compte plusieurs initiatives se concentrant sur le droit d'auteur. Un peu comme le Canada, l'État britannique semble plutôt se concentrer sur les enjeux de droit d'auteur associés à l'utilisation des données, notamment dans les phases d'entraînement des modèles, plutôt qu'à la propriété intellectuelle du contenu généré par l'IA. En effet, à la suite d'une consultation publique sur la propriété intellectuelle, le gouvernement britannique a annoncé vouloir modifier sa loi sur le droit d'auteur pour l'exploration des données (*data mining*), mais pas pour les œuvres et les travaux produits par l'IA générative¹⁸⁰. Aux États-Unis, le US Copyright Office a lancé dans les derniers mois des consultations à propos de l'IA générative, tant par rapport à la portée du droit d'auteur sur les œuvres générées par les modèles d'IA que par rapport à l'incorporation de matériel protégé par le droit d'auteur dans les phases d'entraînements des modèles¹⁸¹.

En terminant, mentionnons la proposition d'amendement du parlement européen concernant plusieurs enjeux en lien avec l'utilisation de l'IA générative, dans le cadre de la proposition du *Artificial Intelligence Act*. En effet, on y suggère que les nouveaux modèles d'IA générative « devraient se conformer à des exigences de transparence supplémentaires, comme le fait de divulguer que du contenu a été généré par l'IA, de concevoir des modèles qui évitent la génération de contenu illégal et de publier de résumés des données protégées par le droit d'auteur utilisées pour l'entraînement »¹⁸².

¹⁷⁹ Lobato et Scarlata (2022).

¹⁸⁰ Gouvernement du Royaume-Uni (2022).

¹⁸¹ Un rapport sera produit dans les prochains mois à partir de ces consultations.

¹⁸² Parlement européen (2023).



Bibliographie de la section Intelligence artificielle, arts et culture

Bjork, C. (2023). *ChatGPT threatens language diversity. More needs to be done to protect our differences in the age of AI.* The Conversation. <https://theconversation.com/chatgpt-threatens-language-diversity-more-needs-to-be-done-to-protect-our-differences-in-the-age-of-ai-198878>.

Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., et Caliskan, A. (2022). *Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale.* <https://doi.org/10.48550/ARXIV.2211.03759>.

Cabinet du ministre de la Culture et des Communications et ministre responsable de la Jeunesse. (2023). *Le ministre Mathieu Lacombe forme un groupe d'experts pour un meilleur rayonnement de la culture québécoise et de la langue française sur les plateformes numériques.*

<https://www.quebec.ca/nouvelles/actualites/details/le-ministre-mathieu-lacombe-mandate-un-groupe-dexperts-pour-un-meilleur-rayonnement-de-la-culture-quebecoise-et-de-la-langue-francaise-sur-les-plateformes-numeriques-47435>

Cetinic, E., et She, J. (2022). *Understanding and Creating Art with AI: Review and Outlook.* ACM Transactions on Multimedia Computing, Communications, and Applications, 18(2), 1-22. <https://doi.org/10.1145/3475799>.

Collier, K. (2023). *Actors vs. AI: Strike brings focus to emerging use of advanced tech.* NBC News. <https://www.nbcnews.com/tech/tech-news/hollywood-actor-sag-aftra-ai-artificial-intelligence-strike-rcna94191>.

Creamer, E. (5 juillet 2023). *Authors file a lawsuit against OpenAI for unlawfully “ingesting” their books.* The Guardian. <https://www.theguardian.com/books/2023/jul/05/authors-file-a-lawsuit-against-openai-for-unlawfully-ingesting-their-books>.

Ferrara, E. (2023). *Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models.* <https://doi.org/10.48550/ARXIV.2304.03738>.

Flick, C., et Worrall, K. (2022). *The Ethics of Creative AI.* In C. Vear et F. Poltronieri (Éds.), *The Language of Creative AI* (p. 73-91). Springer International Publishing. https://doi.org/10.1007/978-3-031-10960-7_5.

Gouvernement du Canada. (2021). *Consultation sur un cadre moderne du droit d'auteur pour l'intelligence artificielle et l'Internet des objets.* Innovation, sciences et développement économique Canada/Innovation, Science and Economic Development Canada.

Gouvernement du Royaume-Uni. (2022). *Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation.* GOV.UK. <https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation>.

Guèvremont, V. (2021). *L'UNESCO et la diversité culturelle.* Revue québécoise de droit international, 165. <https://doi.org/10.7202/1087386ar>.

Guèvremont, V., Saint-Pierre, D., et Brin, C. (2019). *Le rôle de l'État et des médias dans la promotion de la diversité des expressions culturelles à l'ère du numérique : un état des connaissances et des avancées.* Fonds de recherche du Québec. https://frq.gouv.qc.ca/app/uploads/2021/05/cn_rapport_v.guevremont.pdf.



Heikkilä, M. (7 février 2023). *Why detecting AI-generated text is so difficult (and what to do about it)*. MIT Technology Review. <https://www.technologyreview.com/2023/02/07/1067928/why-detecting-ai-generated-text-is-so-difficult-and-what-to-do-about-it/>.

Hong, J.-W., et Curran, N. M. (2019). *Artificial Intelligence, Artists, and Art: Attitudes Toward Artwork Produced by Humans vs. Artificial Intelligence*. ACM Transactions on Multimedia Computing, Communications, and Applications, 15(2s), 1-16. <https://doi.org/10.1145/3326337>.

Jin, D. Y. (2021). *Artificial intelligence in cultural production : Critical perspectives on digital platforms*. Routledge.

La Presse canadienne. (28 avril 2023). *Le projet de loi C-11 pourrait créer un système à deux niveaux, selon des producteurs indépendants*. Le Devoir. <https://www.ledevoir.com/politique/canada/790039/le-projet-de-loi-c-11-pourrait-creer-un-systeme-a-deux-niveaux-selon-des-producteurs-independants>.

Lobato, R. et Scarlata, A. (2022). *Regulating Discoverability in Subscription Video-on-Demand Services*. In: Flew, T., Martin, F.R. (eds) *Digital Platform Regulation*. Palgrave Global Media Policy and Business. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-95220-4_11.

Macek, A. (19 juin 2023). *Copyright and Generative AI*. Slaw. <https://www.slaw.ca/2023/01/19/copyright-and-generative-ai/>.

Mariage, M., et Guèvremont, V. (2022). *La Décennie des langues autochtones (2022-2032) : La Convention sur la protection et la promotion de la diversité des expressions culturelles de l'UNESCO peut contribuer à la préservation et à la revitalisation des langues autochtones*. Minorités linguistiques et société, 18, 235. <https://doi.org/10.7202/1089186ar>.

Ministère de la Culture et des Communications du Québec. (s. d.). *Découvrabilité en ligne des contenus culturels francophones*. <https://www.mcc.gouv.qc.ca/index-i%3D6388.html>. (consulté 6 juillet 2023).

Ministère de la Culture et des Communications du Québec et ministère de la Culture de France. (2020). *Rapport - Mission franco-qubécoise sur la découvrabilité en ligne des contenus culturels francophones*. <https://cdn-contenu.quebec.ca/cdn-contenu/adm/min/culture-communications/publications-adm/rapport/Decouvrabilite-Rapport.pdf>.

Murray, M. D. (2022). *Generative and AI Authored Artworks and Copyright Law*. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4152484>.

Parlement européen. (2023). *AI Act: a step closer to the first rules on Artificial Intelligence*. <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence#:~:text=The%20new%20law%20promotes%20regulatory,that%20significantly%20impact%20their%20rights>.

Parschall, A. (21 avril 2023). *How This AI Image Won a Major Photography Competition*. Scientific American. <https://www.scientificamerican.com/article/how-my-ai-image-won-a-major-photography-competition/>.

Parsons, H. et Cheng, V. (2023). *Canada: Who Owns AI Generated Art? A Primer on Canadian Copyright and AI Artwork*. BLG. <https://www.blg.com/en/insights/2023/02/who-owns-ai-generated-art-a-primer-on-canadian-copyright-and-ai-artwork>.



Radio-Canada. (22 juin 2023). *L'UDA craint que l'IA remplace les artistes de la voix*. Radio-Canada. <https://ici.radio-canada.ca/nouvelle/1990340/uda-doublage-narration-intelligence-artificielle>.

Rioux, M., Mili, H., et Diane-Gabrielle, T. (2021). *Mesure de la découvrabilité des produits musicaux et audiovisuels québécois sur les plateformes numériques*. Fonds de recherche du Québec. https://frq.gouv.qc.ca/app/uploads/2021/10/michele.rioux_decouvrabilite-plateformes-numeriques_rapport-de-recherche.pdf.

Roose, K. (2 septembre 2022). *AI-Generated Art Won a Prize. Artists Aren't Happy*. The New York Times. <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>.

Salkowitz, R. (16 septembre 2022). *AI Is Coming for Commercial Art Jobs. Can It Be Stopped?* Forbes. <https://www.forbes.com/sites/robsalkowitz/2022/09/16/ai-is-coming-for-commercial-art-jobs-can-it-be-stopped/?sh=7ca77c6254b0>.

Small, Z. (10 juillet 2023). *Sarah Silverman Sues OpenAI and Meta Over Copyright Infringement*. The New York Times. <https://www.nytimes.com/2023/07/10/arts/sarah-silverman-lawsuit-openai-meta.html>.

Struppek, L., Hintersdorf, D., Friedrich, F., Brack, M., Schramowski, P., et Kersting, K. (2022). *Exploiting Cultural Biases via Homoglyphs in Text-to-Image Synthesis*. <https://doi.org/10.48550/ARXIV.2209.08891>.

Tchehouali, D., et Agbobli, C. (2020a). *État des lieux de la découvrabilité et de l'accès aux contenus culturels francophones sur Internet*. <https://www.decouvrabilite-francophonie.net/wp-content/uploads/2020/12/Etat-des-lieux.pdf>.

Tchehouali, D., et Agbobli, C. (2020b). *Accessibilité et découvrabilité des contenus francophones : regards croisés*. <https://www.decouvrabilite-francophonie.net/wp-content/uploads/2020/12/Regards-croises.pdf>.

The US Copyright Office. (2023). *Copyright and Artificial Intelligence*. Copyright.gov.

Torrance, A. W., et Tomlinson, B. (2023). *Training Is Everything : Artificial Intelligence, Copyright, and Fair Training*. <https://doi.org/10.48550/ARXIV.2305.03720>.

UNESCO. (1982). *Conférence mondiale sur les politiques culturelles : rapport final*. https://unesdoc.unesco.org/ark:/48223/pf0000052505_fre.

UNESCO. (2005). *La convention de 2005 sur la protection et la promotion de la diversité des expressions culturelles*. https://fr.unesco.org/creativity/sites/creativity/files/2913_16_passport_web_f.pdf.

Zhang, C., Zhang, C., Zhang, M., et Kweon, I. S. (2023). *Text-to-image Diffusion Models in Generative AI : A Survey*. <https://doi.org/10.48550/ARXIV.2303.07909>.

Zhou, K., Lai, Y., et Jiang J. (2022) *VL StereoSet: A study of stereotypical bias in pre-trained vision-language models*. <https://aclanthology.org/2022.aacl-main.40.pdf>.