



< >

Déclaration de Montréal IA responsable_

</ >

PARTIE 6

LES CHANTIERS PRIORITAIRES ET LEURS RECOMMANDATIONS POUR LE DÉVELOPPEMENT RESPONSABLE DE L'IA



CRÉDITS

Ce document est une partie du
**RAPPORT DE LA DÉCLARATION DE MONTRÉAL
POUR UN DÉVELOPPEMENT RESPONSABLE
DE L'INTELLIGENCE ARTIFICIELLE 2018.**
Vous retrouverez le rapport complet [ICI](#).

VERS UNE GOUVERNANCE PARTICIPATIVE DE L'IA

RÉDACTION

Nathalie Voarino, coordonnatrice scientifique,
candidate au doctorat en bioéthique, UdeM

Jean-François Gagné, chercheur au CÉRIUM, UdeM

CONTRIBUTION

Marc-Antoine Dilhac, professeur au Département
de philosophie, UdeM

Christophe Abrassart, professeur à l'École de design
à la Faculté de l'aménagement, UdeM

CHANTIER LITTÉRATIE NUMÉRIQUE

RÉDACTION

Camille Vézy, doctorante en communication, UdeM

CONTRIBUTION

Marie Martel, professeure adjointe à l'École de
bibliothéconomie et des sciences de l'information,
UdeM

Marc-Antoine Dilhac, professeur au Département
de philosophie, UdeM

Dans ce document, l'utilisation du
genre masculin a été adoptée afin de
faciliter la lecture et n'a aucune intention
discriminatoire.

CHANTIER INCLUSION NUMÉRIQUE DE LA DIVERSITÉ

RÉDACTION

Marc-Antoine Dilhac, professeur au Département
de philosophie, UdeM

CONTRIBUTION

Loubna Mekki-Berrada, doctorante
en neuropsychologie, UdeM

Jihane Lamouri, coordonnatrice de la diversité, IVADO

CHANTIER ENVIRONNEMENT

RÉDACTION

Christophe Abrassart, professeur à l'École de design
à la Faculté de l'aménagement, UdeM

CONTRIBUTION

Alessia Zarzani, Ph.D. en aménagement, UdeM
et Ph.D. en Paysage et Environnement,
Université la Sapienza de Roma

Christophe Mondin, professionnel de recherche,
CIRANO

Vincent Mai, doctorant en robotique, UdeM

RECOMMANDATIONS

RÉDACTION

Marc-Antoine Dilhac, professeur au Département
de philosophie, UdeM

Christophe Abrassart, professeur à l'École de design
à la Faculté de l'aménagement, UdeM

Nathalie Voarino, coordonnatrice scientifique,
candidate au doctorat en bioéthique, UdeM

CONTRIBUTION

Les membres du comité scientifique de la
Déclaration de Montréal pour un développement
responsable de l'IA

TABLE DES MATIÈRES

1. INTRODUCTION - Pour une transition numérique créatrice	261
2. VERS UNE GOUVERNANCE PARTICIPATIVE DE L'IA	263
2.1 Comment gouverner les algorithmes : promouvoir la participation citoyenne	263
2.2 Ne pas vivre dans un monde gouverné par les algorithmes : favoriser l'agentivité humaine	267
3. CHANTIER LITTÉRATIE NUMÉRIQUE : Assurer le développement des compétences numériques et la citoyenneté active tout au long de la vie	272
3.1 Outiller les Canadiens de compétences numériques	273
3.1.1 L'écosystème de la littératie numérique	274
Hors du système formel d'éducation et de formation	274
La littératie numérique à l'école	276
3.1.2 La formation professionnelle	276
Développer les compétences liées au numérique dans tous les secteurs	276
Développer les compétences autres que techniques des professionnels en IA	277
3.2 Encourager l'appropriation de la littératie numérique par le renforcement de la citoyenneté active, de la diversité et des solidarités	278
3.2.1 La cybercitoyenneté : compréhension, jugement critique et respect	279
Comprendre, pouvoir agir et critiquer	279
Respecter et responsabiliser	280
Contribuer au bien-être durable de la société	280
3.2.2 L'appropriation de la culture numérique : accessibilité, inclusion et diversité	280
L'inclusion numérique	280

Un enjeu de participation citoyenne	281
Des espaces d'inclusion : les bibliothèques et tiers-lieux	282
4. CHANTIER INCLUSION NUMÉRIQUE DE LA DIVERSITÉ	283
4.1 La neutralité algorithmique en question	285
Des biais humains et des machines impartiales ?	285
Machines à discriminer	286
L'identité biaisée : internet et les SIA	289
4.2 Débiaiser les systèmes d'intelligence artificielle	291
Un problème avec les données	293
Faire parler les algorithmes	294
Représentativité et inclusivité	297
5. CHANTIER ENVIRONNEMENT : IA et transition écologique, enjeux et défis pour une soutenabilité forte.	299
5.1 Transition numérique et transition écologique : une contradiction non résolue	300
5.2 Intelligence artificielle et environnement : défis et opportunités	303
5.2.1 Empreinte environnementale directe et indirecte des systèmes d'intelligence artificielle (SIA)	304
5.2.2 De nouveaux outils prédictifs pour la transition écologique	308
6. LES RECOMMANDATIONS	311
CRÉDITS	I
PARTENAIRES	II

TABLE DES FIGURES ET DES TABLEAUX

Fig. 1 Détail de la couverture du livre de Safiya Umoja Noble, Algorithms of Oppression	291
Fig. 2 Recherche sur le moteur google.com effectuée le 29 octobre 2018	292
Fig. 3 Recherche sur le moteur google.fr effectuée le 29 octobre 2018	292

1. INTRODUCTION - Pour une transition numérique créatrice

La nature perturbatrice (*disruptive*) des technologies du numérique et de l'intelligence artificielle est reconnue unanimement. Mais doit-on voir dans le changement social induit par ces technologies une évolution, une rupture ou plutôt une révolution? C'est une question qui mérite d'être posée, mais qui ne trouvera de réponse que dans quelques décennies. Ce que nous savons aujourd'hui, c'est que ces technologies rendent obsolètes certaines structures de notre organisation sociale et appellent la création de nouvelles structures, qu'elles modifient le marché du travail et le reconfigurent, qu'elles redessinent enfin l'environnement urbain, la mobilité et tous les autres secteurs de la vie sociale.

Posé en ces termes, le problème du changement social rappelle irrésistiblement la thèse de la « destruction créatrice » de l'économiste Joseph Schumpeter. L'idée générale est simple : une innovation technologique offre des possibilités de développement économique et ceux qui s'en saisissent prennent un avantage décisif sur les autres. Une entreprise qui développe ou utilise de nouvelles technologies améliore ainsi son efficacité et peut proposer des produits qui répondent mieux aux besoins des consommateurs ou satisfont de nouveaux besoins. Les entreprises en revanche qui refusent le passage aux nouvelles technologies voient leur existence menacée, et même les plus grandes finissent par disparaître. Les exemples contemporains abondent : Quel adulte né après l'an 2000 sait que des générations ont gardé leurs souvenirs sur des pellicules photographiques qu'il fallait développer avec tout un savoir chimique? En 20 ans, l'industrie de la photographie argentique a été écrasée par les technologies numériques et le nom emblématique de Kodak est désormais relégué à l'histoire des empires industriels. Si le désir de prendre des photographies n'a jamais été aussi fort,

il n'est plus satisfait par l'industrie de l'argentique, ou très marginalement, mais par toute l'industrie numérique de la production de capteurs pour la mise en ligne sur les réseaux sociaux des images.

Nous assistons, avec l'essor des technologies de l'IA, à une nouvelle phase de destruction créatrice, ce « processus de mutation industrielle (...) qui révolutionne incessamment de l'intérieur la structure économique, en détruisant continuellement ses éléments vieillissants et en créant continuellement des éléments neufs. »¹ Aux craintes d'une destruction d'emplois par les SIA, du remplacement des êtres humains et du chômage de masse, certains opposent candidement cette thèse de Schumpeter : s'ils reconnaissent que les SIA remplaceront les êtres humains sur de nombreuses tâches que l'on peut automatiser, les optimistes soutiendront le fait que cela créera d'autres emplois, d'autres besoins et que le marché de l'emploi s'ajustera. La société dans son ensemble s'ajustera, ou plutôt devra s'ajuster :

« Ce processus de Destruction Créatrice constitue la donnée fondamentale du capitalisme : c'est en elle que consiste, en dernière analyse, le capitalisme et toute entreprise capitaliste doit, bon gré mal gré, s'y adapter. »²

Si Schumpeter insiste sur le fait que l'on « doit s'adapter » à ce processus de destruction créatrice, ce « devoir » d'adaptation n'est pas une injonction morale qui répond à un principe éthique, mais un précepte pragmatique. Si une entreprise et une société capitaliste (quel que soit d'ailleurs son régime politique) souhaitent se maintenir, elles doivent s'adapter aux réalités et possibilités offertes

¹ Joseph Schumpeter (1943), *Capitalisme, socialisme et démocratie*, Trad. fr. Gaël Fain, Paris, Payot, 1951, p.128.

² Idem que la précédente.

par les nouvelles technologies. Pourtant, si s'adapter semble une nécessité pour résister à « l'ouragan » technologique (l'image est de Schumpeter), cet ouragan détruira aussi des entreprises et diverses organisations, il marginalisera des villes et des régions, et laissera en arrière des pays entiers qui dépendront d'activités économiques externes. Il peut y avoir de nombreux « perdants » dans cette destruction créatrice, même s'ils font preuve d'une volonté d'adaptation.

En admettant qu'il soit toujours possible de s'adapter – imaginons qu'en 1995 Kodak ait pris la mesure de l'impact du numérique et que la compagnie ait commencé à produire les capteurs qui équipent désormais les appareils numériques – cette adaptation peut prendre beaucoup de temps pour des structures lourdes (usines, grandes entreprises, administrations publiques) alors que le changement technologique peut être très rapide. Dans le cas des nouvelles technologies du numérique et de l'IA, le changement est très rapide et aucune structure sociale n'est adaptée à ce changement : le droit, sans lequel la société devient complètement instable, est beaucoup trop lent à se réformer et à réguler des activités que le législateur peine à comprendre.

Alors quelle sera la part de la destruction dans le développement de l'IA ? Quelle sera la part de la réinvention sociale ? Comment opérer de manière équitable une transformation sociale de l'ampleur de celle qui est induite par le déploiement de l'IA ? Car si l'adaptation aux nouvelles réalités de l'IA est nécessaire, cela ne peut se faire à n'importe quel coût social, ni à n'importe quelle fin. Pour le dire sans détour, les êtres humains ne sont pas très bons pour faire des prédictions et nous ne savons pas quels secteurs seront véritablement touchés par le déploiement de l'IA (les véhicules autonomes peut-être, mais rien n'est certain), ni si l'adaptation à l'IA sera réussie, ni encore quand elle se fera. Dans cette incertitude, il est urgent de trouver des repères pour ouvrir le passage vers une société harmonieuse qui intègre les outils de l'IA.

C'est là tout l'enjeu d'une réflexion sur la transition numérique. Mais pour engager sérieusement cette réflexion, on ne doit pas sombrer dans le pessimisme, ni se faire peur avec des dystopies qui relèvent

de la science-fiction. On s'écartera également de tout optimisme naïf qui voit dans la technologie en général, et dans l'IA en particulier, le remède à tous les maux humains ; les utopies technicistes et scientistes ne nous sont d'aucune utilité. Les utopies politiques nous préservent de la naïveté techniciste, elles indiquent une direction idéale, mais elles ne trouvent aucun ancrage dans le présent et ne permettent donc pas d'enclencher un processus de transformation sociale.

Il convient donc de ne céder ni aux rêves utopiques, ni aux cauchemars dystopiques, mais d'élaborer un réalisme complexe qui prend au sérieux les possibilités offertes par la technologie, qui ne néglige pas les contraintes du présent et ses dynamiques, et qui s'efforce de trouver les leviers d'action pour orienter le déploiement de l'IA vers le bien commun, l'équité sociale et l'agentivité humaine (l'autonomie).

Après avoir mis en œuvre un cadre éthique, nous présentons des réflexions qui ouvrent la voie à une série de recommandations concrètes. Ce travail est le fruit d'un dialogue entre experts, parties prenantes et citoyens. Les ateliers de délibération et de coconstruction de la Déclaration avaient pour objectif explicite de concevoir collectivement des propositions concrètes mettant en place des mécanismes institutionnels pour que le déploiement de l'IA soit socialement responsable et conforme aux principes éthiques de la Déclaration. Les délibérations ont permis de dégager des propositions types et des ordres de priorité dans les actions à mener au cours des prochains mois et des prochaines années. C'est à partir des résultats de ce processus délibératif que nous avons sélectionné des thématiques prioritaires pour outiller les pouvoirs publics, les entreprises et les citoyens, et réaliser une transition numérique créatrice de tissu social, de bien-être collectif, de richesses et de partage : la gouvernance algorithmique ; la littératie numérique ; l'inclusion de la diversité ; la soutenabilité écologique.

Si le monde de l'intelligence artificielle est pour demain, gardons éveillée notre raison pour passer la nuit.

2. VERS UNE GOUVERNANCE PARTICIPATIVE DE L'IA

La gouvernance réfère à une série de politiques et de procédures formelles et informelles. Elle concerne aussi bien des règlements que des lois, des normes et des pratiques, et ce, pour une organisation ou un ensemble d'organisations, qu'elles soient privées ou publiques. La gouvernance algorithmique, elle, renvoie par convention aux procédures qui permettent d'encadrer les dispositifs relatifs à la prise de décision autonome (à des degrés variables) par un système automatisé.

Cependant, une ambiguïté notable est associée à ce terme tantôt référant à « comment gouverner l'intelligence artificielle (IA) » et tantôt à « comment l'IA gouverne ». Cette ambiguïté est soulevée par Musiani (2013) en référence à l'évènement *Governing Algorithms*, qui a eu lieu à New York en mai 2013, dont le titre peut renvoyer tant à la régulation politique des technologies en jeu qu'à un certain pouvoir de gouverner des algorithmes eux-mêmes. Ce dernier renvoie à la question de ce que les algorithmes « peuvent faire », et de quelle manière

ils deviennent des artéfacts de gouvernance par le pouvoir qu'on leur accorde³. Ces deux aspects sont essentiels lorsqu'il est question de la gestion responsable des systèmes d'IA (SIA) dans nos sociétés. Deux principales questions sous-tendent ainsi la gouvernance algorithmique : comment les institutions vont gouverner les algorithmes et à quel point allons-nous vivre dans un monde gouverné par les algorithmes¹ ?

2.1

COMMENT GOUVERNER LES ALGORITHMES : PROMOUVOIR LA PARTICIPATION CITOYENNE

Selon Antoinette Rouvroy et Thomas Berns, la gouvernance algorithmique se déploie en trois temps⁴ :

1. la récolte de quantité massive de données – en particulier par les entreprises privées ;
2. le traitement de ces données et la production de nouvelles connaissances ;
3. l'usage de ces connaissances⁵. Les enjeux de la gouvernance algorithmique sont donc indissociables de ceux des données sur lesquelles les algorithmes apprennent, ou qu'ils analysent. Le grand nombre de données potentialise leur efficacité (lorsqu'il s'agit de leur entraînement), tout comme le poids des décisions qui en émanent.

Des mécanismes et propositions liés à la gouvernance des données ont concrètement vu le jour depuis peu, comme l'entrée en vigueur du règlement général sur la protection des données (RGPD) au sein de l'Union européenne⁶, qui n'est pas sans répercussions à l'international. Des gouvernements, notamment au Québec, rendent accessibles des données publiques sous différentes

³ Francesca Musiani. 2013. «Governance by algorithms». *Internet Policy Review* 2(3).

⁴ Auquel ils préfèrent le terme « gouvernamentalité algorithmique »

⁵ Antoinette Rouvroy et Thomas Berns. 2013. « Gouvernamentalité algorithmique et perspectives d'émancipation ». *Réseaux* (1) : 163-196.

⁶ La Chine possède un équivalent avec le « Personal Information Security Specification » et les États-Unis préfèrent pour le moment une approche marquée par l'absence d'une politique nationale sur les données personnelles.

⁷ World Wide Web Foundation. 2008-2018. « The Open Data Barometer ». En ligne. <https://opendatabarometer.org>

⁸ Ville de Montréal. 2018. « Politique de données ouvertes de la Ville de Montréal ». En ligne. <http://donnees.ville.montreal.qc.ca/portail/politique-de-donnees-ouvertes/>

conditions⁷. La Ville de Montréal développe des politiques sur l'ouverture des données⁸ et sur le logiciel libre⁹ qui vont dans le sens du respect de la vie privée et de la sécurité publique. Des études d'impacts et des analyses de risque fournissent d'intéressants outils aux décideurs¹⁰. Des mécanismes de veille, tels que le « New York City Task Force for Open Data and AI », prennent forme. Le rapport Villani en France prescrit la constitution de « communs de la donnée »¹¹. La stratégie IA du Québec évoque le concept de « data trust », une idée lancée au Royaume-Uni dans un rapport intitulé « Growing the artificial intelligence industry in the UK ». Plus d'une quarantaine de projets à travers le monde vise à impliquer la société civile dans la reformulation du cadre législatif¹². Enfin, certains explorent des techniques permettant d'intégrer la gouvernance des données dans le design même des algorithmes avec un accent sur la représentativité et les genres¹³.

Concernant la production de nouvelles connaissances et leurs usages, c'est la force et la précision des calculs algorithmiques qui seraient à l'origine de la nouvelle forme de pouvoir des SIA¹⁴. Le traitement d'un nombre massif de données (ou *data mining*), rendu possible en quelques secondes, permet l'émergence de corrélations plus ou moins inédites, mais aussi plus ou moins pertinentes. D'une part, en s'appuyant exclusivement sur des données passées, ces analyses peuvent contribuer à informer des outils de gestion ayant pour effet de figer la société dans des paradigmes organisationnels existants (ex. en transport, éducation, justice, santé) et de retarder la mise en place de réformes structurelles parfois nécessaires. D'autre part,

la production automatisée de ces corrélations limite l'intervention humaine et par là même la subjectivité qui y est associée, donnant l'impression d'une objectivité « absolue »⁵. Ces aspects ont été soulevés par les citoyens, lors de la coconstruction, qui craignent les effets déshumanisants d'une approche trop « objective ». Comme le reconnaissent Rouvroy et Berns, cet aspect est problématique seulement si ces corrélations sont utilisées dans le cadre d'interventions politiques et scientifiques sans jamais être remises en question ; en particulier lorsque les décisions qui en découlent affectent les personnes.

Afin de poser des balises quant à l'usage et la production de connaissances algorithmiques, différentes propositions de mécanismes ont vu le jour. Des codes éthiques ont été développés ou sont en cours de développement. L'Institute of Electrical and Electronics Engineers (IEEE)¹⁵ et la conférence Asilomar sur l'IA bénéfique font figure de proue. Des entreprises comme Google, Microsoft ou IBM ont emboîté le pas et rendu publics les principes auxquels elles adhèrent. Ces codes éthiques reposent essentiellement sur l'autorégulation en lien avec la mouvance en matière de responsabilité sociale des entreprises. Des certifications sont en gestation avec un souci de prioriser des modes de corégulation, comme l'initiative de l'International Organization for Standardization (ISO)¹⁶. Cela dit, la majorité des certifications se limitent à des considérations techniques et ne tiennent pas compte des impacts sociaux¹⁷. La stratégie IA Québec suggère en outre de mettre sur pied une organisation mondiale de l'IA responsable. Se développent également des études d'impacts sur l'utilisation

⁹ Ville de Montréal. 2018. « Nouvelle politique au service de l'innovation numérique ». En ligne. <https://beta.montreal.ca/nouvelles/nouvelle-politique-au-service-de-linnovation-numerique>

¹⁰ GovLab. « Open Data's Impact ». En ligne. <http://odimpact.org/>; Ethics & Algorithms Toolkit. « A risk management framework for governments ». En ligne. <http://ethicstoolkit.ai/>

¹¹ Cédric Villani. 2018. « Donner un sens à l'intelligence artificielle : Pour une stratégie nationale et européenne ».

¹² GovLab. *CrowdLaw*. En ligne. <https://crowd.law/> ; LawMaker. 2017. En ligne. <https://lawmaker.io/>

¹³ Christian Sandvig, Kevin Hamilton, Karrie Karahalios, et al. 2014. « Auditing algorithms: Research methods for detecting discrimination on internet platforms ». *Data and discrimination: converting critical concerns into productive inquiry*: 1-23. ; Tolga Bolukbasi et al. 2016. « Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings ». *Advances in Neural Information Processing Systems*: 4349-4357.

¹⁴ Cardon Dominique. 2018. « Le pouvoir des algorithmes ». *Pouvoirs* (1): 63-73.

¹⁵ IEEE. 2018. En ligne. <https://ethicsinaction.ieee.org/>

¹⁶ ISO. 2017. « ISO/IEC JTC 1/SC 42 : Artificial Intelligence ». En ligne. <https://www.iso.org/committee/6794475.html>

¹⁷ Alessandro Mantelero. 2018. « AI and Big Data : A Blueprint for a Human Rights, Social and Ethical Impact Assessment ». *Computer Law & Security Review* 34 (4): 754-772.

de SIA par la fonction publique, telles que celles développées par l'AI NOW Institute, le Conseil du Trésor (Canada)¹⁸ ou encore Nesta en Angleterre. Certains états légifèrent : la Californie oblige par exemple les entreprises en ligne à communiquer publiquement l'usage d'agents conversationnels, afin que l'individu sache s'il a affaire à un humain ou un SIA¹⁹. La gouvernance des algorithmes peut aussi être pensée en termes de conception (design) des algorithmes, notamment par la définition d'objectifs en lien avec le bien-être des individus, par exemple, en introduisant la parité démographique et l'égalité de la probabilité des chances dans l'atteinte des objectifs d'un SIA²⁰.

Une des questions sous-jacentes sur laquelle ont insisté les participants au processus de coconstruction est celle du partage de la responsabilité face à la gestion du développement de l'IA : est-ce aux entreprises ou à l'État de développer ces mécanismes de gouvernance ? L'influence des entreprises propriétaires des algorithmes les plus performants en inquiète plus d'un. Si les conflits d'intérêts potentiels sont dénoncés, ils contestent également la tendance à la marchandisation des données. La position dominante des géants du web, sur des répertoires parfois insoupçonnés de données personnelles conservées sur une longue période, déplaît à plusieurs. En arrière-plan, les questions de la transnationalisation des flux de données et surtout le contrôle des entreprises de la Silicon Valley refont surface. Les études démontrent les conséquences inattendues pour l'individu et la société, dans son ensemble, de l'exploitation des données personnelles à des fins de maximisation du profit

dans un marché oligopolistique²¹. Le rapport de force est asymétrique tant entre les entreprises qu'entre l'entreprise et l'individu ou la société. En effet, concernant notamment les compagnies propriétaires de données massives, certains s'inquiètent de l'apparition de monopoles, renforcés par les fusions de nouveaux fournisseurs de services plus petits²².

Mais s'il faut éviter les situations de monopoles privés, il faut également se garder de favoriser la constitution d'un monopole de l'État sur la production, la propriété, l'accès et l'usage de données, monopole qui n'inspire pas confiance à d'autres participants à la coconstruction. Certaines recherches témoignent de pratiques discutables des États démocratiques à des fins de surveillance et mettent en lumière des partenariats controversés avec le secteur privé en matière de sécurité et de défense²³. Cette relation demande à être clarifiée au-delà de l'aspect stratégique, car elle se déploie dans l'ensemble des champs d'action de l'État. Ni monopole privé, ni monopole d'État : c'est donc la diversité des acteurs qu'il faut préserver.

Au-delà du régime politique, il existe des différences entre les pays en matière de gouvernance des algorithmes²⁴. Cela pose le défi de la coopération internationale et des rivalités entre États qui veulent asseoir leur hégémonie²⁵ normative. Nonobstant les dangers d'abus de pouvoir de part et d'autre, la diversité des modèles nationaux de régulation des données (par exemple ceux des États-Unis, de l'Europe ou de la Chine) provoque des problèmes de coordination à l'international, mais offre aussi des opportunités de dialogue à travers des instances

¹⁸ Treasury Board of Canada Secretariat. 2018. « Responsible Artificial Intelligence in the Government of Canada » Dans Digital Disruption White Paper Series. En ligne. <https://docs.google.com/document/d/1Sn-qBZUXEUG4dVk909eSg5qvfbpNIRhZlefWptBwbxY/edit>

¹⁹ Dave Gershgorn. 2018. « A California law now means chatbots have to disclose they're not human ». Dans Quartz. En ligne. <https://qz.com/1409350/a-new-law-means-californias-bots-have-to-disclose-theyre-not-human/>

²⁰ David Madras, Elliot Creager, Toniann Pitassi et Richard Zemel. 2018. « Learning Adversarially Fair and Transferable Representations ». *arXiv preprint arXiv:1802.06309*.

²¹ Frank Pasquale. 2015. *The Black Box Society. The Secret Algorithms that Control Money and Information*. Harvard University Press.

²² OECD. 2018. « Big data: Bringing competition policy to the digital era ». En ligne. <http://www.oecd.org/competition/big-data-bringing-competition-policy-to-the-digital-era.htm>

²³ Taylor Owen. 2015. « Disruptive Power. The Crisis of the State in the Digital Age », *Oxford Studies in Digital Politics*: 168-188.

²⁴ Allan Dafoe. 2018. « AI Governance: A Research Agenda ». En ligne. <https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI/Agenda.pdf> ; Christoph Bartneck et al. 2007. « The influence of People's Culture and Prior Experiences with Aibo on their Attitudes towards Robots ». *AI & Society* 21 (1-2): 1-14. BCG GAMMA. 2018. « Artificial Intelligence: Have no Fear the Revolution of AI at Work ». En ligne. <https://www.ipsos.com/en/revolution-ai-work>

²⁵ Will Knight. 2018. « China Wants to Shape the Global Future of Artificial Intelligence ». *MIT Technological Review*.

²⁶ Susan Ariel Aaronson et Patrick Leblond. 2018. « Another Digital Divide: The Rise of Data Realms and its Implications for the WTO ». *Journal of International Economic Law* 21: 245-272.

multilatérales²⁶. En ce qui a trait à la gouvernance publique, un encadrement légal et juridique de l'IA s'accompagne de différents risques et suscite des interrogations²⁷ : par exemple, en se centrant trop sur les capacités des dispositifs aux dépens des aspects sociaux de l'automatisation (ce qui peut nuire à la protection des valeurs humaines)²⁸. Est-il possible de réglementer l'IA ? L'État a-t-il véritablement la capacité de le faire ?²⁹.

Le partage de la gouvernance du développement de l'IA entre États et entreprises est sous-tendu par un important dilemme (ressorti des discussions citoyennes, quel que soit le secteur concerné) qui oppose la protection des intérêts individuels à celle des intérêts collectifs. La réponse à ce dilemme constitue un enjeu important qui dépend d'une position normative sur laquelle aucun consensus n'a été observé lors de la coconstruction. Par exemple, ont été soulevées la valeur et l'utilité pour le bien commun, ou bien-être collectif, du partage et de la mise en commun des données (ex. dans un contexte de santé publique, de prévention de la criminalité ou d'éducation), versus la protection de la vie privée au niveau individuel et de la liberté de choisir de partager ou non ses données. Bien qu'on puisse la surmonter, on note une opposition assez classique entre une conception politique qui promeut la liberté individuelle et un espace de non-interférence (protection absolue des données, rejet de toute surveillance) avec une conception qui défend plutôt le bien collectif, l'équité et la transparence des processus, ainsi que des politiques d'allocation des ressources et de partage de renseignements personnels.

En ce qui concerne le monde du travail, ce dilemme a été essentiellement abordé du point de vue de la responsabilité : les participants ont identifié la protection du bien commun selon une

certaine responsabilité collective, défendant qu'il est nécessaire de prendre un virage majeur vers l'économie de partage et que « tous deviennent un peu leurs propres entreprises ». L'autonomie de l'individu dans son parcours de vie et son parcours professionnel (et le bien-être qui y est associé) a été défendue, tout comme le risque de démutualisation et d'une individualisation accrue face aux risques sociaux. À qui doit alors revenir la responsabilité d'assurer le bien-être collectif et individuel lors de la transition numérique ?

Qu'il s'agisse de l'État ou des entreprises, le problème soulevé est celui de la concentration des pouvoirs et d'une verticalité dans leur exercice, au détriment de la représentation de la société civile et d'un partage horizontal du pouvoir d'organiser le déploiement de l'IA. Le contexte actuel se caractérise par quelques joueurs qui dictent les règles sans égard, pour la plupart, aux préférences des citoyens. Si les discussions autour de la gouvernance opposent souvent les institutions publiques aux compagnies privées, une alternative a été proposée lors de la coconstruction : celle d'une gouvernance participative qui donne directement la main aux citoyens en proposant, par exemple, la mise en place d'un espace permanent de concertation. La littérature scientifique démontre la pertinence de la contribution de l'intelligence collective à l'innovation technologique, dont notamment la gouvernance algorithmique³⁰. Si la participation et la collaboration des parties prenantes requièrent du temps, elles n'en sont pas moins sans valeur³¹. L'organisation de « forums hybrides » où collaborent citoyens, experts et administrations sur des objets complexes comme les SIA se justifie en particulier dans un monde incertain où peuvent se déployer à tout moment des controverses socio-techniques et dans lequel aucun acteur ne peut prétendre à l'omniscience³².

²⁷ Matthew U. Scherer. 2015. « Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies ». *Harvard Journal of Law & Technology* 29 (2).

²⁸ Meg Leta Ambrose. 2014. « Regulating the loop: ironies of automation law ». Dans *WeRobot (draft)*. En ligne. <http://robots.law.miami.edu/2014/wp-content/uploads/2014/03/AmbroseWeRobot20141.pdf>

²⁹ J. Danaher. 2015. « Is effective regulation of AI possible? Eight potential regulatory problems » Dans *Philosophical Disquisitions*. En ligne. <http://philosophicaldisquisitions.blogspot.com/2015/07/is-effective-regulation-of-ai-possible.html>

³⁰ Geoff Mulgan. 2017. *Big Mind: How Collective Intelligence Can Change Our World*. Princeton: Princeton University Press. ; John Danaher et al. 2017. « Algorithmic Governance: Developing a Research Agenda through the Power of Collective Intelligence ». *Big Data & Society* 4 (2): 1-27.

³¹ Elizabeth F. Cohen. 2018. *The Political Value of Time*. Cambridge: Cambridge University Press.

³² Michel Callon, Pierre Lascoumes, et Yannick Barthe. 2001. *Agir dans un monde incertain. Essai sur la démocratie technique*. Paris : Le Seuil.

³³ Algorithm Observatory. En ligne. <https://algoritmi.pybossa.com>

Certains tentent ainsi d'ouvrir les algorithmes au public³³. Toutefois, la perception, les préférences et les intérêts des citoyens demeurent dans la grande majorité des cas encore trop peu considérés dans les décisions concernant le déploiement responsable de l'IA.

Dans l'optique de cette gouvernance participative, les citoyens ont souligné l'importance de la contribution des usagers à la conception des outils d'IA et de leur gestion. Cette participation pourrait prendre la forme d'une expérimentation collective centrée sur l'expérience des usagers (*design thinking*) par le biais de prototypes en accès libre (*open source*). Ce matériel accessible à tous constituerait un bien commun numérique (par exemple, les logiciels libres ou les communs de la donnée³⁴) qui semble caractéristique du déploiement du numérique à l'heure actuelle. « Le déploiement du numérique se caractérise par la création de biens publics par les communautés sur internet. Ce processus a supposé l'émergence de formes organisationnelles significativement nouvelles supportées par les technologies de l'information, en particulier les mouvements *open source* puis Web 2.0. »³⁵ Plus qu'une simple forme de propriété, il s'agit ici d'un mode d'organisation coopératif garantissant l'horizontalité des échanges entre pairs, et aussi, la liberté d'expression³⁶. Cette organisation dépend des formes de régulation décidées par les acteurs eux-mêmes²⁷. Ce mode de gouvernance n'est pas lui non plus sans défis, notamment fragile à différentes formes d'*enclosure* (réduction des usages communs) par l'État comme par les compagnies³⁷. Dans une étape ultérieure, on doit envisager que les paramètres sociaux des algorithmes fassent l'objet d'une délibération citoyenne, mieux : d'un codage citoyen. Ce codage ne devrait pas impliquer de compétences supérieures à celles que l'acquisition de la littératie numérique doit garantir, comme nous le verrons dans la prochaine section, et ne nécessite pas non plus la consultation de l'ensemble de la population, mais de groupes multiples de délibération.

Peu importe l'acteur, les participants soulignent la responsabilité collective envers les impacts sociaux de l'IA. Derrière cette idée se cache toutefois une préoccupation : la vitesse du changement technologique laisse peu de temps pour la délibération citoyenne et la réflexion politique. Afin de répondre à ces différents défis, il nous a semblé pertinent de promouvoir une gouvernance qui s'appuie sur la participation citoyenne notamment pour garantir que le déploiement de l'IA se fasse en accord avec les principes et les valeurs fondamentales de notre société. Il apparaît donc indispensable de créer des moyens inclusifs de consultations qui impliquent les citoyens dans toute leur diversité, à différentes étapes du processus de l'encadrement du développement responsable de l'IA (cf. Déclaration de Montréal pour un développement responsable de l'intelligence artificielle, Recommandation 1). Cette participation collective devrait avoir lieu pour la conception des SIA comme pour leur encadrement suite à des retours d'expériences sur leurs dysfonctionnements.

2.2

NE PAS VIVRE DANS UN MONDE GOUVERNÉ PAR LES ALGORITHMES : FAVORISER L'AGENTIVITÉ HUMAINE

Les citoyens ayant participé aux activités de coconstruction soutiennent l'idée d'un certain « humanisme numérique ». Celui-ci implique que les SIA intègrent les principes éthiques ou valeurs humaines fondamentales afin de protéger les intérêts de chacun, incluant notamment le respect de la vie privée, la protection de l'environnement, voire la préservation de ce qui nous définit en tant qu'être humain. Ils craignent une déshumanisation des différents secteurs d'activité touchés par le développement de l'IA, en réduisant les individus à leurs données quantifiables. Ils s'inquiètent également que l'expertise de l'IA soit valorisée

³⁴ Le rapport Villani recommande la constitution de « communs de la donnée », qui inciterait les acteurs économiques à la mutualisation de leurs données et offrirait plus de force aux acteurs publics.

³⁵ Emmanuel Ruzé. 2013. « La constitution et la gouvernance des biens communs numériques ancillaires dans les communautés de l'Internet. Le cas du wiki de la communauté open-source WordPress ». *Management & Avenir* (65):189–205.

³⁶ Hervé Le Crosnier. 2018. « Communs numériques et communs de la connaissance. Introduction. » *tic&société* 12 (1): 1–12.

³⁷ Hervé Le Crosnier. 2011. « Une bonne nouvelle pour la théorie des biens communs ». *Vacarme* 3: 92–94.

au détriment de l'expertise humaine, et qu'il devienne difficile de garder un contrôle sur les algorithmes et leurs décisions. Ces inquiétudes renvoient à la deuxième conception de la gouvernance algorithmique, soit du « comment l'IA nous gouverne ».

Les algorithmes impactent déjà notre vie quotidienne. Différents auteurs signalent l'usage répandu de différentes mesures computationnelles d'évaluation des individus nécessairement approximatives et normatives, ainsi que leurs conséquences potentiellement néfastes et imprévues³⁸. Le danger réside ici dans l'omnipotence du langage informatique qui façonne le monde des possibles sans égard aux subtilités inhérentes au contexte social³⁹. L'usage d'algorithmes marketing qui recommandent des produits ciblés sur la base de l'historique des achats et produits consultés par les individus est un des exemples de l'apparition d'algorithmes qui « gouvernent » en orientant le choix des consommateurs⁴⁰. Les « profils numériques » sont ainsi utilisés, parfois à l'insu des individus, pour différentes fins, au risque de se substituer à leur identité propre²⁸. Ainsi : « Laisser des traces numériques devient synonyme d'une normativité, mais au prix d'une exposition permanente de soi. Ne pas disposer de traces numériques devient a contrario suspect et peut déclencher une surveillance accrue. Il n'est ainsi plus possible d'échapper à l'encerclement des dispositifs électroniques. »²⁸. Le risque d'une mise en danger de l'individu par une désobjectivation est alors souligné⁴¹. Les citoyens défendent cependant que la situation d'une personne ne devrait pas se réduire à des indicateurs quantifiables.

Pour que les algorithmes ne « nous gouvernent » pas, il semble nécessaire, d'une part, de tempérer le pouvoir qu'on leur accorde et, d'autre part, de favoriser un développement des SIA qui va dans le sens de la promotion de **l'agentivité humaine**, soit la capacité d'agir des individus⁴². En effet, considérant la nature de plus en plus autonome des SIA, certains philosophes sont amenés à reconsidérer la notion « d'agentivité morale » jusqu'ici seulement attribuée aux êtres humains⁴³. Cela signifie qu'en « prenant des décisions » les algorithmes se verraient attribuer une forme de responsabilité face aux conséquences des actions issues de leurs recommandations, devenant par là même des « agents » ou acteurs de la société. L'automatisation de l'analyse des données comme de la prise de décision issue de SIA pose en effet d'importantes questions concernant le partage du contrôle entre humains et algorithmes⁴⁴, notamment parce qu'il n'est pas encore possible d'expliquer aux usagers le chemin qui amène un SIA à prendre une décision (la fameuse *black box* de l'IA). Il existe des inquiétudes concernant le déploiement des algorithmes, et leur impact négatif sur le libre arbitre et l'autonomie des individus⁴⁵, qui pourrait potentiellement nuire à la capacité des individus d'assumer certaines responsabilités (soit, nuire à leur agentivité). Les citoyens ont d'ailleurs soulevé un risque de déresponsabilisation et, à terme, de perte de compétences de l'humain en attribuant à l'IA trop de pouvoir, ou en lui reléguant la souveraineté de la décision. Certains ont même soutenu que l'agentivité mériterait d'être un principe à part entière de la Déclaration de Montréal (cf. Partie 7, Les résultats de la coconstruction de l'hiver).

- ³⁸ Jerry Z. Muller. 2018. *Tyranny of the Metrics*. New Jersey: Oxford University Press; Andrea Saltelli et Mario Giampietro. 2017. « What Is Wrong with Evidence Based Policy, and How Can it Be Improved? » *Futures* 91: 62-71; Joshua Newman. 2016. « Deconstructing the Debate over Evidence-Based Policy » . *Critical Policy Studies* 11 (2): 211-226.
- ³⁹ Tarleton Gillepsie, Pablo Bocskowski et Kristen Foot (dir.). 2012. « The Relevance of Algorithms » . *Media Technologies*. Cambridge (MA) : Cambridge University Press; Ed. Finn. 2017. *What Algorithms Want—Imagination in the Age of Computing*. Cambridge (MA): MIT Press.
- ⁴⁰ Fidelia Ibekwe-Sanjuan. 2014. « Big Data, Big machines, Big Science: vers une société sans sujet et sans causalité? » . *XIXe Congrès de la Sfsic. Penser les techniques et les technologies: Apports des Sciences de l'Information et de la Communication et perspectives de recherches* : 1-10.
- ⁴¹ Antoinette Rouvroy et Thomas Berns. 2013. « Gouvernementalité algorithmique et perspectives d'émancipation » . *Réseaux* (1) : 163-196.
- ⁴² Plus spécifiquement, l'agentivité peut référer à l'habileté qu'ont les humains à réfléchir à ce qu'ils valorisent, à déterminer des objectifs et à les réaliser (Isle Oosterlaken. 2015. *Technology and human development*. Routledge, p. 5).
- ⁴³ Merel Noorman. 2012. « Computing and Moral Responsibility » . Dans *The Stanford Encyclopedia of Philosophy*. En ligne. <https://plato.stanford.edu/archives/win2016/entries/computing-responsibility/>
- ⁴⁴ Francesca Musiani. 2013. « Governance by algorithms » . *Internet Policy Review* 2 (3).
- ⁴⁵ Dominique Cardon. 2018. « Le pouvoir des algorithmes » . *Pouvoirs*. 164 (1): 63-73.

Cependant, il est important de souligner que les règles de calcul des algorithmes sont procédurales et non substantielles, c'est-à-dire que les algorithmes n'ont pas de véritable compréhension des informations qu'ils manipulent, ni même des résultats qu'ils produisent³⁷. Ainsi, ce sont bien les humains derrière leur programmation, ceux qui déploient les SIA dans leurs organisations, ou encore ceux qui utilisent leur recommandation, qui doivent être responsables des conséquences d'actions ou de décisions issues de SIA. En d'autres mots, les humains sont les seuls agents de la gouvernance algorithmique, ce sont eux qui doivent prendre la décision finale et être imputables des conséquences néfastes – comme des bénéfiques – issus de l'usage de SIA.

Mais ici s'élève un doute : si les SIA ne gouvernent pas au sens humain, il est tout à fait possible qu'ils soient les agents d'une gouvernance par les procédures, et non par la réflexion, sur la substance éthique et sociale des décisions qu'ils prennent. C'est pourquoi il faut plutôt affirmer normativement, comme l'ont établi les participants aux travaux de la Déclaration de Montréal, que les décisions finales doivent être soumises au contrôle humain, notamment pour les aspects moraux, fonctionnels et politiques de l'IA, malgré (et contre) l'efficacité procédurale de cette dernière. Cette recommandation s'accorde ainsi avec celles de plusieurs rapports internationaux comme celui de la CNIL en France dont le titre est sans équivoque : « Comment permettre à l'homme de garder la main ? »⁴⁶. Une minorité considère comme acceptable de déléguer des microdécisions aux algorithmes selon la gravité des conséquences et la complexité du phénomène. Cette position va dans le même sens que celle des participants qui défendent la nécessité de garder un humain dans la boucle des décisions algorithmiques (*human-in-the-loop*)⁴⁷, d'autant plus importante lorsqu'il s'agit de décisions aux conséquences graves (comme la décision de tuer⁴⁸).

Bien qu'à court et moyen terme, l'humain semble destiné à conserver le contrôle sur l'IA⁴⁹, l'exercice de son agentivité suppose à la fois de préserver certaines compétences et d'assurer l'accès à la connaissance (pour plus de détails, voir la section sur la littératie numérique). En d'autres mots, ceci implique la mise en place d'une gouvernance qui permet l'accès aux compétences et connaissances nécessaires à l'exercice de l'agentivité des individus, mais aussi des organisations qui déploient des SIA et qui doivent garder un rapport réflexif, critique et apprenant sur ces outils.

Une des manifestations de cet exercice en termes de gouvernance est l'obtention du consentement libre et éclairé des personnes qui utilisent des SIA ou font l'objet de leur analyse. Dans cette optique, les citoyens ont défendu qu'il est absolument nécessaire que l'individu sache qui utilise ses données et connaisse les intentions de l'acquéreur, afin de garantir un consentement éclairé. Pour d'autres, un individu devrait avoir accès à une justification compréhensible. Connaître la marge d'erreur de l'option indiquée par un algorithme, et les objectifs qui guident ses recommandations, semble également indispensable pour les citoyens ayant participé à la coconstruction. Cette exigence de transparence n'est pas seulement une condition nécessaire à la confiance, mais un élément clé dans l'exercice de l'agentivité. Dans cette perspective, les organisations devraient selon les citoyens assumer leur responsabilité et prendre les mesures adéquates afin que le « fardeau du consentement » ne repose pas uniquement sur les épaules de l'utilisateur.

Cependant, chez les juristes, le concept de consentement « éclairé » fait couler beaucoup d'encre : il est reçu dans des conditions de plus en plus éloignées de l'esprit du droit⁵⁰. Plus problématique encore pour les urbanistes est l'acquisition des données personnelles sans consentement explicite, notamment dans l'espace public avec la ville intelligente et les objets connectés⁵¹. Concernant le secteur de

⁴⁹ 2015. « AI Timeline Surveys ». Dans *AI Impacts*. En ligne. <https://aiimpacts.org/ai-timeline-surveys/>

⁵⁰ Fred H. Cate et Viktor Mayer-Schönberger. 2013. « Notice and Consent in a World of Big Data ». *International Data Privacy Law* 3 (2): 67-73; Omer Tene et Jules Polonetsky. 2013. « Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property* 11 (5): 239-272.

⁵¹ Rob Kitchin. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Thousand Oak (CA): Sage.

la santé, d'autres questionnent la possibilité d'obtenir, dans les conditions actuelles, un réel consentement éclairé des patients face aux usages de l'IA, notamment en ce qui a trait à la protection de la vie privée et de la confidentialité, mise à mal par la réutilisation exponentielle des données biomédicales⁵². Il semble en effet difficile aujourd'hui de prévoir a priori toutes les utilisations qui seront faites d'un ensemble de données produites, et donc, d'en avertir les individus. Dans ce contexte, l'obligation de revisiter la notion de vie privée au-delà du corpus juridique s'impose⁵³. Certains philosophes introduisent l'idée de droit à l'intériorité⁵⁴ alors que des informaticiens expérimentent, avec des résultats mitigés⁵⁵, des techniques d'anonymisation des données personnelles afin de prévenir la (ré)identification.

Pour bon nombre de chercheurs, l'opacité des réseaux neuronaux constitue précisément le nœud du problème⁵⁶. Et dans le secteur public, l'enjeu est de taille car les algorithmes prennent des décisions qui ont un impact majeur sur la vie quotidienne⁵⁷. Sans explication, surtout en cas d'erreurs ou de dysfonctionnements, et sans recours, les préjudices commis pourraient pénaliser injustement les individus⁵⁸, d'autant plus qu'il n'existe souvent pas de mécanismes de rétroaction visant à réduire les imperfections des systèmes automatisés,

que le calcul reste cryptique et les statistiques dissimulées⁵⁹. C'est ainsi, dans un but de contrôle, que cette transparence est requise, notamment pour assurer une responsabilité humaine face aux abus (et ainsi, les limiter). Par exemple, certaines recherches exposent au grand jour la discrimination générée par de multiples biais inhérents aux SIA. L'une d'elles fait appel à des considérations épistémologiques liées à l'objectivité scientifique : les données sont une construction sociale, un jugement de valeur, elles ne sont pas neutres⁶⁰. Quoique le problème de la fiabilité des données soit amplement documenté dans l'histoire des sciences, les risques de biais prennent des proportions inquiétantes avec l'IA en raison de l'échelle de grandeur : chaque individu est une victime potentielle même si tous ne seront pas affectés⁶¹ (pour plus de détails, voir la section sur l'inclusion numérique de la diversité).

À ce titre, il nous semble donc essentiel de promouvoir et de garantir que le développement des SIA se fasse dans le sens de la préservation, voire de l'augmentation des capacités des personnes et des organisations. Cet aspect fait écho à la Déclaration de la FACIL, qui défend un numérique issu de savoir élaboré en commun et promeut la protection des capacités des citoyens⁶². Dans la même lignée, il est important de citer le mouvement

⁵² Brent Daniel Mittelstadt et Luciano Floridi. 2016. « The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts ». *Science and engineering ethics*. 22 (2): 303–41.

⁵³ Colin J. Bennett et Charles Raab. 2018. « Revisiting the Governance of Privacy: Contemporary Policy Instruments in Global Perspective ». *Regulation & Governance*: 1-18; Neil M. Richards et Jonathan H. King. 2014. « Big Data Ethics ». *Wake Forest Law Review* 49: 393-432.

⁵⁴ Sara Champagne. 2018. « Trois questions sur la vie privée au philosophe Jocelyn Maclure ». *Le Devoir*.

⁵⁵ Groupe de travail « Article 29 » sur la protection des données. 2014. « Avis 05/2014 sur les Techniques d'anonymisation ». En ligne. https://www.cnil.fr/sites/default/files/atoms/files/wp216_fr_0.pdf

⁵⁶ Mike Ananny et Kate Crawford. 2018. « Seeing without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability ». *New Media & Society* 20 (3): 973-989.

⁵⁷ Cathy O'Neil. 2016. *Weapons of Math Destruction. How Big Data Increases Inequality and Threaten Democracy*. New York: Broadway Book.

⁵⁸ ProPublica. 2017. « Machine Bias ». En ligne. <https://www.propublica.org/series/machine-bias>; Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press; Brent Daniel Mittelstadt et al. 2016. « The Ethics of Algorithms: Mapping the Debate ». *Big Data & Society* 3 (2): 1-21.

⁵⁹ Cathy O'Neil. Idem que la précédente.

⁶⁰ Alex Campolo et al. 2017. « AI NOW Report ». *AI Now Institute at New York University*: 15; Luciano Floridi et Mariarosaria Taddeo. 2016. « What is Data Ethics ». *Philosophical Transactions of the Royal Society* 374: 1-5; Erna Ruijter et al. 2018. « Open Data Work: Understanding Open Data Usage from a Practice Lens ». *International Review of Administrative Sciences* 0 (0): 1-17.

⁶¹ Cathy O'Neil. Idem que la précédente.

⁶² FACIL. 2017. « Déclaration des communs numériques ». En ligne. https://wiki.facil.qc.ca/view/D%C3%A9claration_des_communs_num%C3%A9riques

ATM (appropriate technology movement) basé sur l'approche des capacités⁶³ pour réfléchir au développement technologique. Selon ce mouvement, il n'y a aucune raison d'assumer que la technologie la plus avancée est forcément la meilleure option, et la véritable question est la réelle plus-value du développement technologique en ce qui a trait aux capacités humaines. Deux aspects de l'approche par les capacités sont ici particulièrement pertinents. Premièrement, elle implique de se concentrer sur les capacités et le fonctionnement des individus plutôt que sur les seuls moyens (comme, par exemple, les ressources). Deuxièmement, elle implique de porter une attention particulière à la diversité humaine. Le respect de cette diversité est une des principales raisons de centrer les objectifs du développement sur l'expansion des capacités humaines au lieu de l'accès aux ressources. L'atteinte du bien-être est la principale revendication de cette approche. L'agentivité en est un des concepts clés, et suppose que les individus ne sont pas des récepteurs passifs, mais des participants actifs du développement (ici, technologique). Suivant cette idée, les communautés doivent guider le développement technologique (ce qui s'accorde ainsi avec une gouvernance participative) afin qu'il reflète leurs valeurs et objectifs⁶⁴.

Ainsi, dans l'optique de promouvoir la mise en place d'une gouvernance adaptée, il nous a paru nécessaire d'approfondir la réflexion sur trois chantiers prioritaires en vue de formuler des recommandations pour des politiques publiques :

1. **Un chantier sur les enjeux relatifs à la littératie numérique (afin d'assurer le développement des compétences numériques de chacun)**
2. **Un chantier sur les enjeux relatifs à l'inclusion de la diversité**
3. **Un chantier sur l'environnement (afin de garantir un bien-être durable et une soutenabilité écologique forte du développement des SIA).**

Ces chantiers mettent en avant les conditions essentielles (bien que non exhaustives) à la mise en place d'une gouvernance qui se veut en accord avec le bien-être des individus dans toutes leur diversité et la promotion de leur agentivité, notamment dans le cadre d'une gouvernance participative. Ces conditions nous paraissent essentielles afin d'assurer un impact positif des algorithmes sur la vie des individus, et que chacun puisse être acteur de sa réalité numérique dans l'optique d'une responsabilité collective.

⁶³ L'approche des capacités est issue des travaux de Amartya Sen et Martha Nussbaum. « Ces deux penseurs défendent que l'évaluation du progrès en matière de développement ne devrait pas être faite en termes de revenus ou de ressources, mais en termes de capacités humaines individuelles - ou de ce que les gens sont réellement capables de faire et d'être » (Isle Oosterlaken. 2015. *Technology and human development*. Routledge, p. 2, traduction libre). Ainsi, une capacité peut se comprendre comme la capacité à réaliser un bien humain fondamental comme celui de se déplacer, d'être en santé ou encore de développer sa pensée.

⁶⁴ Isle Oosterlaken. 2015. *Technology and human development*. Routledge, p. 2

3. CHANTIER LITTÉRATIE NUMÉRIQUE : Assurer le développement des compétences numériques et la citoyenneté active tout au long de la vie

*Déclaration de Montréal IA responsable,
Principe 2.4 :*

« Il est indispensable d'encapaciter les citoyennes et les citoyens face aux technologies du numérique en assurant l'accès à différents types de savoir, le développement de compétences structurantes (la littératie numérique et médiatique) et la formation de la pensée critique. »

La littératie numérique est reconnue par des organismes tels que l'UNESCO et l'OCDE comme étant **au cœur de la participation et de l'engagement social et citoyen dans une société de l'information et une économie de la connaissance**. Elle est ainsi définie comme étant « l'habilité à accéder, gérer, comprendre, intégrer, communiquer, évaluer et créer de l'information de façon sécuritaire et appropriée par les outils numériques et les technologies en réseaux pour la participation à la vie économique et sociale »⁶⁵. Elle inclut des compétences évoquées également sous les termes de littératie des technologies de l'information et la communication, littératie de l'information,

littératie des données et littératie médiatique⁶⁶. La littératie numérique ne se résume donc pas seulement au fait de savoir utiliser des outils technologiques, elle inclut également une dimension critique amenant à savoir prendre des décisions éclairées quant à cette utilisation.

Dans une société de l'information qui relève avant tout d'une civilisation de l'écrit, la littératie numérique dépend de l'aptitude à comprendre et utiliser l'information écrite dans la vie courante (littératie et alphabétisme fonctionnel). Elle s'inscrit ainsi dans un continuum allant de la littératie de base à la capacité de comprendre et interagir de façon éclairée avec des SIA.

LITTÉRATIE

LITTÉRATIE NUMÉRIQUE

LITTÉRATIE DE L'IA



⁶⁵ GAML, UNESCO Institute for Statistics. 2018. « A draft report on a global framework on digital literacy skills for indicator 4.4.2 : Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills ». En ligne. <http://gaml.cite.hku.hk/wp-content/uploads/2018/03/DLGF-draft-report-for-online-consultation-all-gaml.pdf> (Traduction libre, p.3)

⁶⁶ Idem que la précédente

Lors des délibérations citoyennes de la Déclaration de Montréal, l'enjeu de la littératie numérique a été abordé dans tous les secteurs. Les citoyens ont souligné qu'il était **nécessaire de former la population aux pratiques et enjeux** de l'intelligence artificielle. Cette formation permet d'acquérir les **compétences à la fois techniques et critiques** indispensables pour que tout individu puisse agir de façon autonome, éclairée et responsable en tant que **travailleur et citoyen** dans une société en transition. Les objectifs principaux sont ainsi de favoriser le **développement d'une bonne compréhension et d'un sens critique** par rapport au fonctionnement des systèmes d'intelligence artificielle (SIA), à leur utilisation et aux nouvelles normes qui leur sont liées, notamment en matière de données personnelles. La littératie numérique s'est donc imposée aux citoyens comme un ensemble de compétences pour entretenir notamment une **vigilance collective afin de développer et utiliser des SIA de façon responsable**.

Si les jeunes sont ainsi visés par la littératie numérique dès les classes de primaire, celle-ci s'adresse également aux étudiants, quelle que soit leur spécialisation, mais aussi aux professionnels de tout secteur (santé, éducation, justice, ressources humaines et administration publique en particulier). Les concepteurs et programmeurs de SIA sont par ailleurs également concernés par la littératie numérique, notamment au nom de la nécessité d'« intégrer la formation en éthique liée aux technologies et aux enjeux de l'IA dans le cursus des ingénieurs et dans la formation continue » (Mémoire de l'Ordre des ingénieurs du Québec, Recommandation 5).

Dans cette perspective, les principales pistes de solution proposées au fil du processus de coconstruction de la Déclaration ont été de développer la littératie numérique à tous les âges de la vie, par une éducation à la fois technique et éthique. Celle-ci se ferait via des canaux formels tels que l'école, l'université ou la formation professionnelle continue, mais également via des « formations populaires » à l'IA (cf. Partie 3, Rapport des résultats des ateliers de coconstruction de l'hiver, section 5.2) et aux réalités numériques qui lui sont liées afin de rejoindre l'ensemble de la population canadienne.

Les citoyens ont par ailleurs soulevé deux enjeux de justice sociale liés à la littératie numérique : celle-ci doit se développer de façon accessible à tous, sur l'ensemble du territoire canadien, elle doit également se développer de manière à maintenir une diversité de profils d'apprenants et en portant attention aux différentes formes d'intelligence. Cela nécessite donc de penser à des solutions pour que le développement de la littératie numérique soit structurellement accessible, inclusif, porteur et facteur de diversité.

À la suite de cette réflexion issue des délibérations citoyennes, nous proposons d'explorer en deux temps le développement de la littératie numérique afin de présenter des recommandations dans la lignée des principes de la Déclaration de Montréal, en particulier ceux d'autonomie, de responsabilité, d'équité, de diversité et de solidarité. L'objectif principal est d'assurer le développement des compétences numériques tout au long de la vie, que ce soit par des canaux formels (école, université, formation professionnelle) ou des canaux informels (hors de ces systèmes). Ce développement de la littératie numérique comme apprentissage tout au long de la vie a lui-même deux visées :

1. **développer le capital humain des Canadiens en les outillant de compétences numériques ;**
2. **encourager l'appropriation de la littératie numérique par le renforcement de la citoyenneté active, de la diversité et de la collaboration entre membres d'une communauté, favorisant ainsi le développement d'une société apprenante.**

3.1

OUTILLER LES CANADIENS DE COMPÉTENCES NUMÉRIQUES

Les compétences numériques correspondent à la capacité de repérer, comprendre, organiser, évaluer, créer et diffuser de l'information par l'intermédiaire de technologies numériques ; elles permettent de réaliser des objectifs liés à l'apprentissage, au travail et à la participation sociale. Le renforcement de ces compétences numériques représente un enjeu

d'innovation et de développement économique à l'échelle du Canada qui mise sur le développement des compétences des Canadiens afin d'accéder plus facilement à des emplois bien rémunérés et de faire grandir la classe moyenne, comme en témoigne le *Plan pour l'innovation et les compétences*⁶⁷.

L'approche du capital humain⁶⁸ semble être ainsi privilégiée : il s'agit d'investir dans les compétences et savoirs que les individus peuvent acquérir pour favoriser la croissance économique et la compétitivité internationale en formant une main-d'œuvre compétente. Cela se confirme, entre autres, par des investissements du ministère de l'Innovation, des Sciences et du Développement économique du Canada (ISDE) pour le développement d'initiatives de littératie numérique, mais également par la stratégie pancanadienne en matière d'intelligence artificielle pilotée par le Canadian Institute For Advanced Research (CIFAR), ainsi que par des stratégies nationales sur la main-d'œuvre telle que celle du ministère du Travail, de l'Emploi et de la Solidarité sociale (TESS) du Québec pour assurer la transition numérique.

Dans un contexte de transition, la littératie numérique se présente ainsi, dans un premier temps, sous l'angle des compétences qu'elle permet d'acquérir pour accéder à des emplois et/ou assurer la transformation des emplois actuels. Des mesures garantissant l'égalité d'accès au développement de ces compétences, et une égalité des opportunités d'accéder à ces emplois, gagnent cependant à être mises de l'avant.

Ces compétences numériques peuvent se distinguer en trois types, combinant savoirs technologiques et jugement critique⁶⁹ :

1. **Les compétences numériques de base, dont tout individu a besoin pour participer aux sociétés contemporaines. Il peut s'agir de comprendre comment chercher de l'information fiable (littératie médiatique ou de l'information), communiquer avec d'autres individus de façon bienveillante et sécuritaire, apprendre à utiliser**

des données (littératie des données), ou encore se servir de différents logiciels et applications, et ainsi savoir interagir avec confiance avec la technologie.

2. **Les compétences propres à un secteur de travail spécifique dont les métiers sont amenés à se transformer en demandant d'interagir davantage avec des SIA qu'il s'agit donc d'apprendre à utiliser, et ce de manière responsable.**
3. **Les compétences des professionnels du numérique qui représentent l'ensemble des compétences nécessaires pour développer de nouvelles technologies, de nouveaux services et produits. Cela inclut par exemple la maîtrise de différents langages de programmation, de méthodes d'analyse de données ou encore de techniques d'apprentissage automatique.**

Dans une perspective d'apprentissage tout au long de la vie, ces compétences sont amenées à être développées à la fois dans les systèmes formels de l'école, de l'université et de la formation professionnelle, mais également de plus en plus hors de ces systèmes, par des initiatives d'organisations privées ou sans but lucratif. Un équilibre est à trouver pour encourager les maillages entre entreprises de technologies éducatives, OBNL, écoles et universités pour que l'éducation au numérique se développe comme un bien public accessible à tous.

3.1.1 L'écosystème de la littératie numérique

HORS DU SYSTÈME FORMEL D'ÉDUCATION ET DE FORMATION

Le Canada compte déjà de nombreux programmes d'éducation et de formation pour le développement de la littératie numérique. De **nombreuses organisations hors du système éducatif formel** se développent et proposent une grande variété d'activités.

⁶⁷ Canada - Ministère des Finances. 2017. « Bâtir une classe moyenne forte. Chapitre 1 : Compétences, innovation et emplois pour la classe moyenne ». *Ottawa: Ministère des Finances*. En ligne. <https://www.budget.gc.ca/2017/docs/plan/budget-2017-fr.pdf> (p.48-52).

⁶⁸ Theodore W. Schultz. 1961. « Investment in human capital ». *The American Economic Review* 51 (1): 1-17; Gary Stanley Becker. 1975. *Human capital: A theoretical and empirical analysis with special reference to education*. Chicago, IL: University of Chicago Press.

⁶⁹ TAnnalise Huynh et al. 2018. « Levelling Up: The Quest for Digital Literacy ». Dans *Brookfield Institute for Innovation and Entrepreneurship*. En ligne. <http://www.deslibris.ca/ID/10097218> (p. 4-5)

Le ministère de l'Innovation, des Sciences et du Développement économique du Canada (ISDE) a lancé **deux programmes d'envergure pour le développement d'initiatives de littératie numérique** : **CodeCan** (50 millions de dollars investis sur une période de deux ans à compter de 2017-2018) et **le Programme d'échange en matière de littératie numérique (PELN)** (29,5 millions de dollars investis de 2018 à 2022).

Les initiatives financées par CodeCan encouragent les possibilités de formation en programmation et en perfectionnement des **compétences numériques** chez les jeunes Canadiens de la maternelle à la fin du secondaire⁷⁰. Le programme finance également la formation et le perfectionnement professionnel des nouveaux enseignants par l'intermédiaire d'HabiloMédias qui crée plusieurs ressources en ligne⁷¹. Le PELN quant à lui finance des projets destinés à un plus large public afin de « doter les Canadiens des compétences nécessaires pour utiliser les ordinateurs, les appareils mobiles et internet efficacement et en toute sécurité »⁷².

Les approches des organisations **hors du système éducatif formel** sont variées – elles rassemblent du mentorat, des formations payantes, des programmes dans des centres communautaires, des ateliers dans des bibliothèques, des cours en ligne – et s'adressent à plusieurs publics, des jeunes aux seniors, en passant par les étudiants post-secondaires et les professionnels. Les activités consistent entre autres en des formations intensives (bootcamps) pour l'apprentissage de différents langages de programmation (ex : Lighthouse Labs, Canada Learning Code), des ateliers techno-créatifs dans les fab labs (Communautique) et des bibliothèques (TechnoCultureClub) pour apprendre l'impression 3D par exemple, des compétitions de création d'applications mobiles pour encourager l'entrepreneuriat technologique chez les jeunes

filles (Technovation Montréal), des ressources en ligne sur la littératie numérique pour les parents, enfants et enseignants (HabiloMédias), et de nombreuses autres⁷³. Le développement des cours en ligne (MOOC) permet également de valider des connaissances ou simplement nourrir la curiosité de façon autonome. Plusieurs de ces initiatives sont financées par des subventions fédérales ou provinciales (cf. le PELN et CodeCan), mais également par des investissements privés. C'est le cas par exemple d'Ubisoft qui investit plus de 8 millions de dollars dans le programme CODEX regroupant des « initiatives à tous les niveaux de scolarité qui positionnent le jeu vidéo comme source de motivation et moteur d'apprentissage pour le développement de la relève techno-créative au Québec »⁷⁴.

Si l'offre de formations et d'activités éducatives hors du système formel est riche et variée, celle-ci n'est pas clairement organisée et il peut être difficile de s'orienter vers celle qui correspond le mieux à nos besoins selon l'âge, le niveau de connaissance, les intérêts. Soulignons cependant l'existence de quelques outils qui facilitent l'orientation soit par le mentorat en ligne (Academos), soit par le recensement des activités pour développer des compétences numériques (Ma Vie Techno).

Une meilleure structuration de cet écosystème bénéficierait aux individus cherchant à se former en matière de numérique à tout âge de leur vie, aux acteurs du milieu (startups, petites ou moyennes entreprises, OBNL, centres communautaires, etc.) qui pourraient davantage partager leurs pratiques, mais également aux décideurs dont les choix pourraient être facilités en ayant un meilleur aperçu des réalités et des besoins des acteurs qui participent à la mise en place de l'école et l'université de demain et rendent possible l'apprentissage tout au long de la vie⁷⁵.

⁷⁰ Gouvernement du Canada. 2018. « Initiatives financées par CodeCan ». En ligne. <https://www.ic.gc.ca/eic/site/121.nsf/fra/00003.html>

⁷¹ Habilo medias. En ligne. <http://habilomedias.ca/ressources-pedagogiques>

⁷² Gouvernement du Canada- Ministère de l'Innovation, des Sciences et du Développement économique. 2018. « Programme d'échange en matière de littératie numérique ». En ligne. <http://www.ic.gc.ca/eic/site/102.nsf/fra/accueil>

⁷³ Ce rapport fait un panorama très riche des organismes et types d'activités offertes sur le territoire canadien : Annalise Huynh et al. 2018. « Levelling Up: The Quest for Digital Literacy ». Dans *Brookfield Institute for Innovation and Entrepreneurship*. En ligne. <http://www.deslibris.ca/ID/10097218>

⁷⁴ Ubisoft. « Codex ». En ligne. <https://montreal.ubisoft.com/fr/programme-codex/>

⁷⁵ Cela pourrait s'inspirer de l'observatoire de la EdTech en France qui « rassemble les acteurs du numérique pour l'éducation et la formation » : Observatoire de la EdTech. En ligne. <http://www.observatoire-edtech.com>

LA LITTÉRATIE NUMÉRIQUE À L'ÉCOLE

L'éducation au numérique se fait de plus en plus par les **canaux formels**, à l'école primaire et secondaire, et dans les instituts post-secondaires, par le biais de nouveaux programmes et l'implémentation de la technologie comme outil d'apprentissage.

Au Québec, la littératie numérique ne figure pas encore en tant que telle dans le *Programme de formation de l'école québécoise*. Elle se rapproche cependant de l'étude des médias qui constitue un domaine général de formation (comme la santé, l'entrepreneuriat, la citoyenneté et l'environnement), mais ne représente pas une discipline au même titre que le français, les mathématiques, les arts plastiques ou l'histoire et la géographie⁷⁶. Le *Plan d'action numérique en éducation et en enseignement supérieur*⁷⁷ du ministère de l'Éducation et de l'Enseignement supérieur (MEES) annonce cependant trois orientations (et 33 mesures) pour soutenir le développement de l'éducation au et par le numérique :

Orientation 1 : Soutenir le développement des compétences numériques des jeunes et des adultes.

Orientation 2 : Exploiter le numérique comme vecteur de valeur ajoutée dans les pratiques d'enseignement et d'apprentissage.

Orientation 3 : Créer un environnement propice au déploiement du numérique dans l'ensemble du système éducatif.

À ce jour, la formation en matière de littératie numérique est cependant dispensée de manière aléatoire, sans évaluation et souvent à l'initiative des professeurs et des directions, que ce soit au primaire et secondaire, au collégial ou à l'université. Plusieurs initiatives existent pour structurer la formation de

compétences numériques, que ce soient celles des élèves et étudiants comme celles des enseignants et professeurs. C'est le cas par exemple du REPTIC⁷⁸ qui met en place des activités et établit un profil des habiletés informationnelles, cognitives, méthodologiques et technologiques, ou encore de l'Association of College & Research Libraries (ACRL) qui a créé un modèle pour la littératie de l'information en éducation supérieure⁷⁹. Ce type d'initiatives gagnent à être clairement intégrées dans la politique éducative afin d'avoir davantage d'impact et de contribuer à la structuration de la formation en matière de littératie numérique.

3.1.2 La formation professionnelle

DÉVELOPPER LES COMPÉTENCES LIÉES AU NUMÉRIQUE DANS TOUS LES SECTEURS

En matière de formation professionnelle, le développement des compétences numériques est mis de l'avant, notamment dans la *Stratégie nationale sur la main-d'œuvre 2018-2023*⁸⁰ du ministère du Travail, Emploi et Solidarité sociale (TESS) du Québec, afin d'« accroître la productivité sur le marché du travail par la formation continue »⁸¹. Tout travailleur est visé, qu'il occupe ou non un emploi.

Les personnes sans emploi pourront s'adresser à Services Québec, aux établissements de formation, aux organismes spécialisés en développement de l'employabilité et aux entreprises d'entraînement qui « collaboreront pour identifier les besoins de formation et d'apprentissage, développer l'offre de formation, intégrer les compétences numériques dans l'aide à la recherche d'emploi et préparer adéquatement la main-d'œuvre à l'acquisition de compétences numériques. »⁸² Les personnes

⁷⁶ HabiloMédias. 2016. « Québec - Aperçu de l'éducation aux médias ». En ligne. <http://habilomedias.ca/ressources-pedagogiques/resultats-dapprentissage-en-education-aux-medias-et-litteratie-numerique-par-province-et-territoire/quebec-aperçu-de-léducation-aux-médias>

⁷⁷ Québec - Ministère de l'Éducation et de l'Enseignement supérieur. 2018. « Plan d'action numérique en éducation et enseignement supérieur ». En ligne. http://www.education.gouv.qc.ca/fileadmin/site_web/documents/ministere/PAN_Plan_action_VF.pdf

⁷⁸ Réseau des répondantes et répondants TIC. 2002-2018. En ligne. <https://www.reptic.qc.ca/>

⁷⁹ ACRL. 2015. « Framework for Information Literacy for Higher Education ». En ligne. <http://www.ala.org/acrl/standards/ilframework>; version francophone : PDCI de l'Université du Québec. 2015. « Référentiel de compétences informationnelles en enseignement supérieur ». En ligne. <http://ptc.quebec.ca/pdci/referentiel-de-competences-informationnelles-en-enseignement-superieur>

⁸⁰ Québec: Ministère Travail, Emploi et Solidarité sociale. 2018. « Stratégie nationale sur la main d'œuvre 2018-2023. Le Québec à l'ère du plein emploi ». En ligne. https://www.mtess.gouv.qc.ca/publications/pdf/Strat-nationale_mo.PDF

⁸¹ Titre de l'axe 3.3 de la *Stratégie sur la main d'œuvre 2018-2023*

⁸² Mesure 41 de la *Stratégie sur la main-d'œuvre 2018-2023*, p. 70

occupant déjà un emploi qui auraient besoin de développer ou d'actualiser leurs compétences numériques pourront faire appel à Emploi Québec qui « accroîtra ses achats de formations à temps partiel en fonction des besoins définis dans les régions du Québec »⁸³. La mise à niveau des travailleurs en matière de compétences numériques fait donc partie de la stratégie du TESS, mais remarquons qu'il n'est pas encore fait mention de la nécessité d'adaptation de la main-d'œuvre à la multiplication de SIA et de systèmes automatisés amenant la transformation de plusieurs métiers.

La formation continue doit par ailleurs être également proposée et prise en charge par les employeurs, en particulier quand les métiers de leurs employés sont amenés à se transformer par l'utilisation de SIA pour différentes tâches, comme c'est le cas dans la santé, l'éducation, la justice, les administrations publiques et privées. De telles formations doivent alors non seulement **permettre d'acquérir les compétences techniques pour savoir utiliser des SIA dans des tâches quotidiennes**, mais elles doivent également **amener ces professionnels utilisateurs de SIA à en faire un usage responsable** en étant sensibilisés aux dimensions éthiques et sociales de cette utilisation. Ces formations pourraient ainsi mettre l'accent sur la prise de décision assistée par SIA de sorte que l'intervention humaine ne soit pas exclue (cf. principe de responsabilité) – en particulier quand la décision affecte la vie, la qualité de la vie ou la réputation d'une personne – et que la mesure des implications éthiques et sociales de la décision soit toujours prise en considération et devienne une habitude professionnelle.

Dans cette perspective, des **codes de déontologie** (cf. Partie 4, Rapport des résultats des ateliers de la coconstruction de l'hiver, section 5.2) ou bien une forme de « **permis d'utiliser les algorithmes et l'IA** »⁸⁴ dans des secteurs particuliers (santé, marketing, ressources humaines, justice, éducation, administration publique) pourraient être créés et obtenus **après avoir suivi des modules de formations spécifiques offerts par des universités et écoles**

spécialisées. Tous les professionnels interagissant avec des outils d'aide à la décision de type SIA devraient également recevoir la **formation adéquate leur permettant de faire un usage responsable de ces outils et de pouvoir justifier leur décision** (cf. principe de participation démocratique).

DÉVELOPPER LES COMPÉTENCES AUTRES QUE TECHNIQUES DES PROFESSIONNELS EN IA

La formation des compétences en IA fait l'objet de nombreux financements en éducation supérieure, notamment par le biais du Canadian Institute for Advanced Research (CIFAR). Celui-ci est chargé d'opérationnaliser la **stratégie pancanadienne en matière d'intelligence artificielle** qui vise à maintenir et développer l'excellence en recherche du Canada⁸⁵ par quatre grands axes :

1. **l'attraction et la rétention de talents en intelligence artificielle,**
2. **la collaboration entre les pôles scientifiques sur le territoire (Edmonton, Montréal, Toronto),**
3. **le développement d'un leadership de pensée sur les implications économiques, éthiques, politiques et juridiques de l'IA et**
4. **le soutien d'une communauté de recherche nationale**¹⁶.

Plus de la moitié du budget (86,5 millions de dollars) est dédiée à la création de chaires de recherche en intelligence artificielle afin d'attirer et retenir les meilleurs chercheurs universitaires dans les domaines de l'apprentissage profond et de l'apprentissage par renforcement. Tandis que ces chaires semblent relever exclusivement du domaine de l'informatique, un programme IA et société est également annoncé pour financer des groupes travaillant sur les implications politiques et économiques de l'intelligence artificielle afin d'informer le public et les politiciens sur ces enjeux.

⁸³ Idem que la précédente

⁸⁴ CNIL. 2017. « Comment permettre à l'Homme de garder la main ? Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle ». En ligne. <https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de> (p.55)

⁸⁵ CIFAR. 2017. «Pan-Canadian Artificial Intelligence Strategy Overview». En ligne. <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>

Le financement de la création de connaissances en IA inclut donc la réflexion éthique, politique, économique et sociale de l'IA. Celle-ci gagne à être transmise auprès des étudiants et des chercheurs en IA afin qu'ils intègrent ces enjeux dans leurs pratiques de développement de l'IA. Des initiatives émergent en ce sens, comme le défi de l'informatique responsable lancé par la fondation Mozilla pour explorer de nouvelles façons d'enseigner l'éthique aux étudiants en informatique⁸⁶. Mieux formés aux enjeux éthiques et sociaux des SIA et des systèmes d'acquisition et d'archivage des données personnelles (SAAD) qu'ils créent ou utilisent, et sensibilisés à la part de responsabilité qu'ils ont dans leur développement, les concepteurs et programmeurs pourraient choisir d'employer, ou non, certains algorithmes et dispositifs d'IA de façon plus éclairée quant à leurs effets potentiels⁸⁷.

3.2

ENCOURAGER L'APPROPRIATION DE LA LITTÉRATIE NUMÉRIQUE PAR LE RENFORCEMENT DE LA CITOYENNETÉ ACTIVE, DE LA DIVERSITÉ ET DES SOLIDARITÉS

La formation des compétences numériques tout au long de la vie, qu'il s'agisse de compétences de base ou de compétences professionnelles, demande ainsi d'articuler apprentissages techniques et sensibilisation pour une utilisation éclairée et une

conduite socialement responsable. La littératie numérique inclut donc la littératie des données, la littératie médiatique ainsi qu'une littératie de l'intelligence artificielle qui comprend l'analyse et l'évaluation critique des enjeux des SIA. Elle n'est pas seulement un enjeu de développement économique par le renforcement du capital humain de chacun, mais également un enjeu éducatif et humaniste⁸⁸ qui vise à encourager la citoyenneté active dans l'espace numérique.

En intégrant la littératie numérique par la dynamique de l'apprentissage tout au long de la vie (ATLV), on souligne les valeurs humanistes et démocratiques d'inclusion et d'émancipation sur lesquelles l'ATLV repose selon l'UNESCO :

« Face aux enjeux et aux défis mondiaux de l'éducation, l'apprentissage tout au long de la vie, 'du berceau au tombeau', est une philosophie, un cadre de pensée et un principe d'organisation de toutes les formes d'éducation, basé sur des valeurs humanistes et démocratiques d'inclusion et d'émancipation ; il a un caractère global et fait partie intégrante de la vision d'une société fondée sur le savoir »⁸⁹.

⁸⁶ Mozilla. « Responsible Computer Science Challenge ». En ligne. <https://foundation.mozilla.org/en/initiatives/responsible-cs/> ; Fast Company. 2018. « Mozilla's ambitious plan to teach coders not to be evil ». En ligne. <https://www.fastcompany.com/90248074/mozillas-ambitious-plan-to-teach-ethics-in-the-age-of-evil-tech>

⁸⁷ Cf. les rapports *Portrait 2018 des recommandations internationales en éthique de l'IA* (en particulier, ce qui a trait au rapport de la Royal Society) ; *Rapport de la coconstruction en ligne et des mémoires reçus* (en particulier, le mémoire de l'Ordre des ingénieurs du Québec, celui du AI Ethics meetup et les réponses au questionnaire en ligne).

⁸⁸ Dans la lignée de Kapil Dev Regmi. 2015. « Lifelong learning: Foundational models, underlying assumptions and critiques ». *International Review of Education* 61 (2): 133-151.

⁸⁹ UNESCO. 2009. « Cadre d'action de Belém. Exploiter le pouvoir et le potentiel de l'apprentissage et de l'éducation des adultes pour un avenir viable ». En ligne. <http://uil.unesco.org/fileadmin/keydocuments/AdultEducation/fr/Cadre%20d%27action%20de%20Bel%C3%A9m.pdf>

La littératie numérique fait ainsi partie de ces savoirs permettant à chacun d'acquérir les connaissances et compétences nécessaires pour réaliser ses aspirations et contribuer à une société⁹⁰ dont le numérique fait de plus en plus intimement partie. Comprise comme un enjeu d'épanouissement personnel et collectif, celle-ci doit se développer de manière accessible, inclusive et renforçant les solidarités de citoyens actifs dans une société apprenante. Face au discours prônant le développement des compétences numériques au nom d'un impératif d'employabilité, la littératie numérique gagne à se développer de manière à favoriser une diversité des intelligences, des profils, des genres et des générations, de manière à ralentir une certaine uniformisation de la société en entretenant sa diversité.

3.2.1 La cybercitoyenneté : compréhension, jugement critique et respect

La notion de « cybercitoyenneté » renvoie à l'exercice de ses droits fondamentaux, de ses compétences politiques, comme la participation aux débats et aux décisions publiques, et de ses devoirs de civilité dans l'univers numérique. Un cybercitoyen développe ou utilise des moyens numériques pour participer à la vie politique. Il peut aussi se définir comme membre d'une communauté numérique qui agit politiquement.

Cette notion soulève cinq grands enjeux : la liberté d'expression et la qualité de l'information, la responsabilité individuelle et sociale des acteurs du numérique, la transparence, le respect de la vie privée, et la justice.⁹¹

COMPRENDRE, POUVOIR AGIR ET CRITIQUER

La cybercitoyenneté relève des principes de respect de l'autonomie, de responsabilité, mais également de participation démocratique et de protection de l'intimité et de la vie privée. Elle invite en effet à développer dès le plus jeune âge la **capacité de comprendre l'écosystème numérique, en particulier celui des SIA, et d'acquérir des savoir-faire pour naviguer dans l'information, protéger nos outils et données personnelles, partager du contenu, etc.** Cette compréhension permet de forger un **consentement** qui est véritablement libre et éclairé, elle permet également de pouvoir **contester** des décisions algorithmiques et, éventuellement, de **vérifier** la pertinence des paramètres et des données pris en compte dans cette décision quand celle-ci est justifiée de façon intelligible. Dans cette perspective, la littératie numérique outille pour comprendre le numérique et les décisions algorithmiques, et donne également la capacité d'agir dans cet univers, face à ces décisions.

Pour cela, la formation d'un **jugement critique** est nécessaire non seulement pour savoir utiliser des outils numériques et des SIA de façon responsable, mais également pour **savoir accorder de la confiance ou douter** de certaines sources, de certaines recommandations et incitations – voire défier certaines formes de manipulation ou de domination. En intégrant la formation de ce jugement critique, la littératie numérique devrait permettre aux individus de faire preuve de davantage de liberté dans leur utilisation de SIA, en évitant de se faire imposer un mode de vie particulier (cf. principe d'autonomie).

⁹⁰ À partir de : UNESCO. 2015. « Forum mondial sur l'éducation, 19-22 mai 2015 », Incheon, République de Corée, cité dans : Baril. 2017. « L'apprentissage tout au long de la vie : définition, évolution, effets sur la société québécoise ». 9^e Journée professionnelle de Bibliothèque et Archives nationales du Québec, Montréal. En ligne. http://www.banq.qc.ca/documents/services/espace_professionnel/milieux_doc/services/journees_professionnelles/apprentissage/Baril.pdf

⁹¹ Québec: Commission de l'éthique en science et en technologie (CEST). 2018. « Éthique et cyber-citoyenneté: Un regard posé sur les jeunes ». En ligne. http://www.ethique.gouv.qc.ca/fr/assets/documents/CEST-Jeunesse/CEST-J-2017/CEST_avis_Cybercitoyennete_FR_vf_Web.pdf (p.1)

RESPECTER ET RESPONSABILISER

En combinant compréhension et jugement critique, la littératie numérique devrait ainsi amener chacun à se responsabiliser quant à la protection de son intimité et de celle des autres (principe de vie privée) – sans toutefois que les autres acteurs voient leur responsabilité diminuer quant au respect de la vie privée et de l'autonomie des utilisateurs d'outils numériques et de SIA. Il peut s'agir de protéger ses données personnelles, décider de contribuer à les partager et demander à les vérifier. Cela peut être également de savoir se comporter de manière respectueuse envers ou via des SIA, en évitant d'adopter un comportement de harcèlement ou de cyberintimidation par l'intermédiaire de médias numériques. L'espace numérique est un espace de vie collective, la littératie numérique doit permettre d'améliorer le vivre-ensemble dans cet espace, tout en invitant les gouvernements, les entreprises, les écoles et les parents à assumer leur part de « responsabilité en matière d'éducation, de sensibilisation et d'autonomisation (...) dans un souci de cohérence et en fonction des valeurs de notre société »⁹².

Cette combinaison de compréhension, de jugement critique et de respect permet d'outiller des personnes capables de faire respecter leurs libertés d'utilisateurs et de citoyens, de participer avec bienveillance à une société comptant de plus en plus d'agents artificiels, et liée par des médias numériques, mais aussi de faire entendre leur voix quant au développement de SIA.

CONTRIBUER AU BIEN-ÊTRE DURABLE DE LA SOCIÉTÉ

La littératie numérique peut par ailleurs aider dans la réponse aux enjeux de santé mentale – tels que des troubles anxieux, des troubles de l'humeur et les problèmes de dépendance⁹³, et de développement

durable associés au développement des SIA (principe de bien-être).

En matière de santé mentale, le développement de la littératie numérique devrait se faire dès le plus jeune âge en limitant le recours à du matériel numérique de façon à limiter le risque de dépendance. L'enseignement des fondements de la culture algorithmique devrait ainsi se faire le plus possible par des outils et techniques non numériques⁹⁴. L'éducation au numérique gagnerait ainsi à transmettre des façons de préserver des moments de déconnexion, à encourager l'imagination et à gérer, voire réduire, les facteurs de stress et d'anxiété générés par des interactions numériques.

L'apprentissage de pratiques environnementales responsables mérite également de faire partie intégrante des enseignements de littératie numérique. Cela pourrait par exemple consister à sensibiliser la population aux coûts énergétiques des SIA. Cela pourrait également concerner l'acquisition de compétences créatives et de réflexes de bricolage pour réparer des objets plutôt que de les jeter, et ainsi limiter les déchets numériques.

3.2.2 L'appropriation de la culture numérique : accessibilité, inclusion et diversité

L'INCLUSION NUMÉRIQUE

Le développement de la littératie numérique se heurte à l'enjeu de fracture numérique qui évoque l'existence d'une « inégalité face aux possibilités d'accéder et de contribuer à l'information, à la connaissance et aux réseaux, ainsi que de bénéficier des capacités majeures de développement offertes par les technologies de l'information et de la communication »⁹⁵. Cette fracture peut se creuser selon l'accessibilité aux infrastructures numériques

⁹² CEST. « Responsabilité individuelle et sociale des acteurs du numérique », p.33. Idem que la précédente.

⁹³ Secrétariat à la jeunesse-Québec. 2018. « Agir sur les problèmes de santé mentale ». En ligne. <https://www.jeunes.gouv.qc.ca/politique/habitudes-vie/sante-mentale.asp>

⁹⁴ CNIL. 2017. « Comment permettre à l'homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle ». En ligne. https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf (p. 54)

⁹⁵ Michel Élie. 2001. « Le fossé numérique, l'internet facteur de nouvelles inégalités ? ». Problèmes politiques et sociaux (861) : 33-38. Cité dans : Québec: Commission de l'éthique en science et en technologie (CEST). 2018. « Éthique et cyber-citoyenneté: Un regard posé sur les jeunes ». En ligne. http://www.ethique.gouv.qc.ca/fr/assets/documents/CEST-Jeunesse/CEST-J-2017/CEST_avis_Cybercitoyennete_FR_vf_Web.pdf p. 14)

(équipement), et selon la capacité à développer les compétences et connaissances nécessaires pour utiliser pleinement ces technologies. La littératie numérique devrait se développer de telle sorte **que le numérique soit un outil d'inclusion**, utilisable par toute personne, quel que soit son genre, son âge, son handicap, sa situation géographique.

Étant donné que le territoire canadien est inégalement équipé en infrastructures pour offrir à tous les Canadiens un accès internet à haut débit, de même que les écoles, bibliothèques et autres espaces communautaires qui sont eux aussi inégalement équipés en technologie, la littératie numérique au Canada souffre d'une inégale répartition sur le territoire. Cet état de fait amène à exiger des politiques publiques et des programmes qui se donnent pour but de réduire la « fracture numérique » (géographique et générationnelle) et l'écart entre ceux qui ont des compétences numériques et ceux qui ont un faible niveau de littératie numérique.

Dans cette optique, une table de concertation intersectorielle et interrégionale en littératie numérique au Québec a été lancée par le Printemps numérique en septembre 2018 pour identifier « des priorités d'actions collectives afin d'améliorer la qualité et les conditions d'intervention en matière de littératie numérique »⁹⁶. Cette table de concertation s'inscrit dans le cadre du projet Jeunesse QC 2030 soutenu par le Secrétariat à la jeunesse du Québec pour connaître les réalités des jeunes Québécois face au numérique en allant à leur rencontre à l'occasion de cafés numériques dans différentes villes du territoire québécois⁹⁷.

L'inclusion numérique peut également être favorisée par une éducation au numérique faite de façon à contribuer au développement des solidarités entre les personnes, les communautés et les générations (cf. principe de solidarité). L'apprentissage intergénérationnel et par les pairs gagne ainsi à être valorisé.

UN ENJEU DE PARTICIPATION CITOYENNE

En étant indissociable d'une formation à la cybercitoyenneté, la littératie numérique relève d'une responsabilité partagée permettant à chacun, sur l'ensemble du territoire, de participer à la vie collective dont le numérique devient une partie intégrante. Si la participation citoyenne en venait à être sollicitée dès la phase de conception de certains SIA pour délibérer sur les paramètres sociaux des SIA, leurs objectifs et les limites de leurs décisions (cf. principe de publicité), tout individu pourrait ainsi être inclus dans cette discussion et ainsi prendre part à la recherche de solutions créatives, éthiquement acceptables et socialement responsables (cf. principe d'autonomie).

La littératie numérique serait en même temps indissociable d'une culture numérique en prenant la forme d'une éducation populaire par des initiatives de médiation auprès de toutes les catégories de population à travers l'ensemble du territoire⁹⁸. Cela a été proposé tant par les citoyens de la Déclaration de Montréal (cf. Partie 4, Rapport des résultats des ateliers de coconstruction de l'hiver, section 5.2) que dans les rapports comme celui de la CNIL ou encore l'IEEE qui souligne l'importance d'une sensibilisation publique aux questions d'éthique et de sécurité liées aux technologies d'intelligence artificielle, à la fois pour assurer une utilisation éclairée et sécuritaire, mais également pour diminuer la peur, la confusion et l'ignorance à propos des enjeux que posent ces technologies.

⁹⁶ Printemps numérique. 2018. « Pour une égalité des chances face au numérique ». En ligne. <https://mailchi.mp/358e547609f8/le-pn-lance-la-premiere-table-de-concertation-en-littratie-numrique-au-qubec?e=d4a8cb83f8>

⁹⁷ Secrétariat à la jeunesse Québec. 2018. « Jeunesse QC 2030 ». En ligne. <http://www.printempsnumerique.ca/projets/projet/jeunesse-qc-2030/>

⁹⁸ CNIL. 2017. « Comment permettre à l'Homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle ». En ligne. https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf CEST-J-2017/CEST_avis_Cybercitoyennete_FR_vf_Web.pdf p. 14)

DES ESPACES D'INCLUSION : LES BIBLIOTHÈQUES ET TIERS-LIEUX

Les bibliothèques jouent un rôle clé dans l'inclusion et la littératie numériques, que ce soit par l'accès à des technologies, aux informations en ligne de qualité liées à la santé, l'éducation, l'emploi ou par le renforcement de compétences numériques critiques dans une perspective d'apprentissage tout au long de la vie. On peut alors parler d'encapacitation (*empowerment*) numérique, ou de développement des capacités qui permettent de vivre, apprendre et travailler dans une société numérique.

L'inclusion numérique est reliée à la littératie numérique, car elle met l'accent sur les politiques, les services et les espaces qui visent à réduire les barrières à l'accès, faciliter le partage des savoirs (notamment locaux ou critiques) et la participation active des publics exclus en les priorisant. En ce sens, l'encapacitation numérique est une condition de l'inclusion numérique dans le contexte d'émergence des SIA.

Les bibliothèques qui intègrent des approches encapacitantes et inclusives en matière d'accès, de formation, d'espace de participation active et sécuritaire (« *safe space* ») — tant pour l'intégrité physique que l'exercice de la liberté d'expression — sont désignées comme tiers-lieux.

Les tiers-lieux, qu'ils soient bibliothèques, fab labs⁹⁹, centres communautaires ou culturels, favorisent la confiance et l'engagement par le biais d'espaces communs, ouverts, flexibles qui facilitent les usages collectifs, voire la conception collaborative, les apprentissages en communautés numériques, les conversations démocratiques transformatrices. Le « faire ensemble » à travers la création de liens sociaux et de communs amplifie l'inclusion et la littératie numérique en contribuant à une citoyenneté active, créatrice à terme de « vivre-ensemble ».

⁹⁹ Ou « laboratoires de fabrication ». Ce sont des lieux dédiés à la fabrication de projets via un ensemble de logiciels et solutions libres et open-source. FabFoundation. *Fab Lab*. 2018. En ligne. <http://fabfoundation.org/index.php/what-is-a-fab-lab/index.html>

4. CHANTIER INCLUSION NUMÉRIQUE DE LA DIVERSITÉ

Si les désaccords sur le sens de la démocratie sont encore vifs, un idéal démocratique fait pourtant consensus : l'inclusion de tous dans la société des égaux. Inversement, l'exclusion d'une partie de la population de la communauté politique pour des raisons économiques, sociales, politiques, culturelles, religieuses ou encore ethniques, entre autres, apparaît comme un échec démocratique si cette exclusion n'est pas intentionnelle, et comme une faute politique si elle résulte de discriminations intentionnelles. L'idéal de la démocratie, quels que soient ses échecs de fait, et peut-être même en raison de ses défaillances à les surmonter, est contenu dans cette formule : personne ne doit être laissé pour compte ; *no one should be left behind*.

Comme on pouvait s'y attendre, les citoyens qui ont participé aux ateliers délibératifs de la Déclaration ont affirmé avec force cet idéal d'inclusion et se sont inquiétés que le développement de l'IA se fasse au détriment d'une partie de la population, aggrave les inégalités ou engendre de nouvelles discriminations, de manière directe ou indirecte et insidieuse¹⁰⁰. Le problème des discriminations et l'enjeu de l'inclusion ont été abordés à partir des principes de justice et de démocratie, mais aussi de connaissance et de vie privée. Si le principe de justice suffit à justifier l'importance de l'inclusion de la diversité et en fait une finalité démocratique, il existe aussi une raison instrumentale : la diversité peut être recherchée comme un moyen pour améliorer la réflexion collective de façon à stimuler la créativité et l'innovation. L'homogénéisation de la société et de ses parties (élites économiques, classe politique, chercheurs, employés de bureau, etc.) conduit le plus souvent, sinon toujours, à une perte de créativité et de capacité à s'adapter aux changements technologiques et sociaux.

Les délibérations ont permis d'affiner la compréhension des enjeux de l'inclusion démocratique dans le développement de l'IA et ont contribué à enrichir les principes de la Déclaration, faisant apparaître la pertinence de formuler un principe d'inclusion de la diversité qui ne se réduit pas à celui de la participation démocratique ni de l'équité, mais qui leur est étroitement lié.

¹⁰⁰ Cf. *Rapport des résultats des ateliers de coconstruction de l'hiver*, « Les grandes catégories de risques et enjeux du développement responsable de l'IA », section « Justice sociale ».

7. PRINCIPE D'INCLUSION DE LA DIVERSITÉ

Le développement et l'utilisation de SIA doivent être compatibles avec le maintien de la diversité sociale et culturelle et ne doivent pas restreindre l'éventail des choix de vie et des expériences personnelles.

Ce principe d'inclusion de la diversité appliquée aux systèmes d'intelligence artificielle (SIA) rappelle le droit à l'égalité et à la non-discrimination proclamé par la Déclaration universelle des droits de l'homme (art. 7)¹⁰¹ et par les différentes chartes des droits et constitutions des sociétés démocratiques.

L'article 10 de la Charte des droits et libertés de la personne, au Québec, développe le lien entre égalité, liberté et droit à ne pas être discriminé ; il mérite d'être cité intégralement :

« Toute personne a droit à la reconnaissance et à l'exercice, en pleine égalité, des droits et libertés de la personne, sans distinction, exclusion ou préférence fondées sur la race, la couleur, le sexe, l'identité ou l'expression de genre, la grossesse, l'orientation sexuelle, l'état civil, l'âge sauf dans la mesure prévue par la loi, la religion, les convictions politiques, la langue, l'origine ethnique ou nationale, la condition sociale, le handicap ou l'utilisation d'un moyen pour pallier ce handicap.

Il y a discrimination lorsqu'une telle distinction, exclusion ou préférence a pour effet de détruire ou de compromettre ce droit. »¹⁰²

Enfin, selon l'article 15 de la Charte canadienne des droits et des libertés :

« La loi ne fait acception de personne et s'applique également à tous, et tous ont droit à la même protection et au même bénéfice de la loi, indépendamment de toute discrimination, notamment des discriminations fondées sur la race, l'origine nationale ou ethnique, la couleur, la religion, le sexe, l'âge ou les déficiences mentales ou physiques. »¹⁰³

Si ces différents principes éthiques et juridiques sont partagés par l'ensemble des participants aux délibérations du processus de coconstruction de la Déclaration qu'ils soient citoyens, experts ou parties prenantes, et par les différents acteurs du développement de l'IA, le passage à des recommandations et des actions respectant ces normes éthiques et juridiques de haut niveau n'est pas évident et pose une série de difficultés. La première réside dans le repérage des discriminations et des exclusions qui seraient liées à l'utilisation de SIA. Une deuxième difficulté consiste à identifier les potentielles causes de discrimination, puis à cerner les conséquences de la discrimination sur l'autonomie des personnes, sur leur capacité à mener une vie digne et conforme à leur conception du bien. Une autre difficulté porte sur la compréhension de la diversité, et on peut la résumer de la manière suivante : Diversité de quoi ? Inclusion dans quoi ? Nous ne donnerons pas de définition a priori et trop restrictive de la diversité. Le processus de

¹⁰¹ CNIL. 2017. « Comment permettre à l'Homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle ». En ligne. https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf (p. 54)

¹⁰² Charte des droits et libertés de la personne, RLRQ c C-12, <http://canlii.ca/t/69v6g>, chapitre I.1., art. 10, consulté le 2018-11-13.

¹⁰³ Loi de 1982 sur le Canada. 1982. ch. 11 (R.-U.), art. 15.

coconstruction a permis d'aborder différents aspects de la diversité qui sont souvent étudiés séparément : diversité des résultats produits par les SIA, diversité des données qui alimentent les SIA, diversité de leurs utilisateurs, diversité des sexes (genre et sexualité) et des minorités culturelles dans le développement des SIA, etc.

Parmi les acquis du processus de coconstruction qu'il faut souligner, on note l'idée que les SIA façonnent le contexte de formation de notre identité, en réduisant la diversité des options disponibles et en procédant par stéréotypes, et affectent ainsi profondément notre identité même. Le deuxième acquis est que l'enjeu de la diversité ne doit pas être compris seulement du point de vue du fonctionnement des SIA, mais du point de vue des mécanismes sociaux qui rendent possibles son développement et son déploiement. Il s'agit d'une perspective de « critique sociale ». Dit plus simplement, les milieux de recherche en informatique et de conception industrielle des SIA, entre autres, sont des lieux qui n'échappent pas à la reproduction des discriminations sexuelle, sociale, culturelle et ethnique, et peuvent même contribuer à les rendre encore plus vives. Ces discriminations, comme nous le noterons plus loin, sont rarement intentionnelles, mais plutôt indirectes, systémiques et non recherchées. Elles n'en sont pas moins très problématiques, et reflètent des mécanismes plus profonds et cachés d'exclusion ou de marginalisation.

Un enjeu que le processus de coconstruction n'a permis que d'effleurer, mais qu'on ne saurait négliger, est celui de l'inclusion de la diversité dans le déploiement de l'IA au niveau international. On ne peut ignorer que le développement de l'IA est un enjeu stratégique et économique important et qu'il fait l'objet d'une concurrence internationale intense dans laquelle certaines nations sont structurellement désavantagées et sont perçues comme des espaces de prédation (main-d'œuvre informatique bon marché, données non protégées, faillites des services publics de santé, de justice et de police, ressources naturelles déjà contrôlées par des compagnies étrangères).

4.1

LA NEUTRALITÉ ALGORITHMIQUE EN QUESTION

Des biais humains et des machines impartiales ?

Dès que l'on aborde le fonctionnement des SIA et leur intérêt social, on bute sur un paradoxe : l'intérêt des algorithmes (apprenants ou non) est qu'ils permettent de parvenir automatiquement au résultat visé en éliminant les erreurs de raisonnement des êtres humains. Or l'idée que les algorithmes puissent également amplifier les biais humains n'est pas sans fondement et tempère la confiance que l'on a dans l'impartialité algorithmique. Pour bien comprendre le paradoxe, il faut revenir d'abord à l'hypothèse selon laquelle les algorithmes, en particulier ceux des SIA, sont moins biaisés que les humains.

La première chose à considérer est que les êtres humains, quoique doués d'une intelligence qui dépasse en complexité celle des algorithmes, sont prompts à faire des erreurs dues à leur état émotionnel¹⁰⁴, à leur niveau de fatigue, à leurs soucis, mais surtout à des biais cognitifs et idéologiques difficiles à éliminer. Les biais cognitifs sont des modes intuitifs de pensée qui déforment (biaisent) le raisonnement logique et induisent des croyances erronées¹⁰⁵. Parmi la quarantaine de biais recensés, relevons le biais de confirmation qui est la tendance à ne rechercher que les informations confirmant nos croyances et à refuser celles qui les contredisent. Un biais qui a un rôle important dans la formation de biais idéologiques et dans la genèse des exclusions sociales directes est le biais de négativité selon lequel on retient davantage les expériences négatives que les expériences positives (ce biais permet aussi d'apprendre d'erreurs tragiques). Les êtres humains ont tendance à ignorer leurs propres biais et ne pas les voir à l'œuvre dans leur raisonnement rapide. C'est particulièrement problématique dans les cas où il faut prendre dans l'urgence une décision qui a des conséquences importantes sur soi et autrui.

¹⁰⁴ Sur les différentes dimensions des émotions dans les processus de connaissance et de raisonnement, cf. Joseph Ledoux. 1998. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York : Simon & Schuster. Voir aussi les travaux de Antonio R. Damasio. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace & Company.

¹⁰⁵ Sur les biais cognitifs, cf. Daniel Kahneman et Patrick Egan. 2011. *Thinking, Fast and Slow*, New York: Farrar, Straus & Giroux.

L'utilisation d'algorithmes pour résoudre des problèmes ou pour prendre la meilleure décision dans des cas d'urgence, d'information incomplète et d'incertitude, s'avère précieuse. Dans son sens le plus fondamental, un algorithme est un ensemble d'instructions, une recette construite en étapes programmables, développée dans le but d'organiser et d'agir sur un corpus de données, pour accomplir rapidement un résultat escompté¹⁰⁶. L'intérêt de leur conception et de leur utilisation est double : l'algorithme permet d'automatiser une tâche et de parvenir toujours au résultat voulu ; il permet d'éliminer les biais qui affectent le raisonnement des êtres humains. Un des célèbres cas qui a permis de réduire la mortalité infantile à la naissance est le test du Dr Apgar qui consiste en une formule avec 5 variables (battements cardiaques, respiration, réflexes, tonus musculaire et couleur) pour évaluer l'état de santé du nouveau-né¹⁰⁷. Avec une procédure très rudimentaire, la formule du Dr Apgar a permis de faire mieux que l'intuition humaine dans des circonstances de jugement difficile. C'est le principe du triage dans les services d'urgence dans les hôpitaux.

Kahneman (2011) nous convainc aisément que les algorithmes sont généralement plus fiables que les humains parce qu'ils ne sont pas biaisés. Bien sûr, ce sont les êtres humains qui conçoivent l'algorithme en fonction du résultat qu'ils recherchent. Mais l'utilisateur de l'algorithme n'a plus qu'à l'appliquer pour obtenir le bon résultat. Dans le cas des SIA, la machine embarque un algorithme apprenant capable d'identifier des motifs dans des ensembles gigantesques de données, d'apprendre d'elle-même en interagissant avec le milieu et d'appliquer différentes lignes d'instruction. Débarrassés des biais qui faussent les raisonnements humains, les SIA sont censés être des instruments neutres qui donnent des résultats neutres.

Les citoyens soutiennent à ce sujet des affirmations en apparence contradictoires. D'un côté, ils attendent que les SIA soient plus neutres ou impartiaux que les êtres humains, et forment l'espoir que des juges numériques rendent de meilleurs jugements. D'un autre côté, ils s'en méfient, mettant en doute leur impartialité. C'est l'inquiétude qu'ils manifestent dans les domaines de la justice et de la police prédictives, mais aussi dans le secteur de la santé et des ressources humaines. Sous le vernis de la neutralité, la prise de décision automatique pourrait dissimuler des biais et exacerber, voire créer des discriminations.

Machines à discriminer

Si l'on peut nourrir des craintes à l'égard des SIA, il n'est pas facile de démontrer s'ils sont biaisés, de dire lesquels le sont, ni quelles en sont les causes. Dans le processus de consultation de la Déclaration, les participants avaient un scénario écrit d'avance pour susciter leur réflexion. Les biais algorithmiques et les discriminations qui en découlent étaient clairement identifiables. Hors de ce contexte, il n'est pas évident de dégager les discriminations ou les effets de marginalisation engendrés par les algorithmes et encore moins de les corrélés à des biais algorithmiques. Une analyse critique du fonctionnement des SIA et un suivi des trajectoires socio-économiques des populations et des individus vulnérables permettent néanmoins dans un premier temps de dégager des corrélations entre l'utilisation des SIA et certaines discriminations¹⁰⁸.

Les récents travaux de Virginia Eubanks¹⁰⁹ ont permis de documenter précisément les discriminations algorithmiques. Dans un livre au titre évocateur, *Automating Inequality* (automatiser l'inégalité), Eubanks a étudié avec rigueur les systèmes

¹⁰⁶ Benjamin Peters (ed.). 2016. *Digital Keywords: A Vocabulary of Information Society and Culture*. Princeton : Princeton University Press. Version préliminaire accessible en ligne : Benjamin Peters (ed.). 2016. « Digital Keywords : A Vocabulary of Information, Society and Culture ». En ligne. <http://culturedigitally.org/wp-content/uploads/2016/07/Gillespie-2016-Algorithm-Digital-Keywords-Peters-ed.pdf>

¹⁰⁷ Daniel Kahneman et Patrick Egan. 2011. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux, chap. 21; Atul Gawande. 2010. *A Checklist Manifesto*. Penguin Books India.

¹⁰⁸ Voir le rapport: The Citizen Lab - University of Toronto. 2018. « Bots at the Gate ». En ligne. <https://ihrp.law.utoronto.ca/sites/default/files/media/IHRP-Automated-Systems-Report-Web.pdf> (p.31)

¹⁰⁹ Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press

automatiques qui jugent quelles personnes sont éligibles à des prestations sociales, à des remboursements médicaux, et lesquelles ne sont plus éligibles. L'éligibilité peut être déterminée par un ensemble de critères qui comprend les ressources financières actuelles, les données sur l'habitat et la zone de résidence, l'état de santé, etc. Avec l'arrivée des ordinateurs, les bases de données se sont agrandies et les administrations publiques comme les compagnies privées (banques, assurances) y ont accès et peuvent traiter des données historiques : la personne a-t-elle des antécédents médicaux ? Depuis quand ? Combien de fois ? A-t-elle toujours remboursé à temps son crédit ? Avec le développement des SIA, non seulement on peut traiter beaucoup plus de données qui affinent les profils des administrés ou des clients, mais on peut faire des prédictions sur leurs comportements, leur solvabilité ou encore sur l'évolution de leur santé. En effet, l'une des vertus des SIA, qui expliquent en partie leur déploiement massif dans les administrations et les compagnies privées, est cette capacité à faire des prédictions de plus en plus riches et souvent très précises. L'une des raisons de leurs succès est que les êtres humains sont assez prévisibles dans leurs comportements et que les motifs dans leurs habitudes sont facilement décelables pour un SIA bien conçu.

Mais ce que cette fonction de prédiction rend possible, c'est un profilage des personnes dans le but d'éviter toute prise de risque qui entraîne un coût pour l'administration ou la compagnie privée. Dès qu'un algorithme signale un risque lié au profil d'une personne, il enclenche aussi des processus de surveillance plus étroite ou d'exclusion des programmes d'aide sociale, d'assurance santé, de recrutement, etc.

Les simples systèmes de notation (score) qui étaient le principe même de la formule du Dr Apgar permettant de sauver des vies, tendent ainsi

à automatiser l'exclusion et les inégalités en signalant systématiquement les personnes pauvres ou en situation de vulnérabilité comme des personnes à risque. Comme le montre Virginia Eubanks, ces systèmes automatiques ont tendance à punir les pauvres et les personnes marginalisées. De fait, en les signalant comme des personnes à risque, les SIA les exposent à des risques supplémentaires de marginalisation¹¹⁰. Ces outils de prédiction, par boucle de rétroaction, sont susceptibles de créer ainsi les difficultés qu'elles prétendent signaler¹¹¹. Par exemple, un système automatique de recrutement par notation des candidats à une entrevue d'embauche apprendra à rejeter les candidatures des personnes qui présentent un risque d'absentéisme, ou de plus faible performance au travail, parce qu'elles résident loin de leur futur lieu de travail. Or ce type de décision qui discrimine les candidats en fonction de leur lieu peut renforcer des inégalités socio-économiques : c'est exactement ce qui s'est produit dans le cas de la compagnie Xerox documenté par Cathy O'Neil¹¹². Les personnes dont la candidature était rejetée habitaient dans des zones résidentielles éloignées... et pauvres. Avec des notes plus faibles en raison d'un contexte économiquement défavorable, ces personnes ont moins de chance de trouver un emploi et connaissent plus de risques de précarité. Dans le cas de Xerox, la compagnie s'est aperçue de ce résultat discriminatoire et a modifié le modèle de l'algorithme : « *The company sacrificed a bit of efficiency for fairness* »¹¹³.

Les cas problématiques se multiplient : le calcul prédictif semble reproduire ou accentuer des inégalités et des discriminations en cours dans la société. L'algorithme de la compagnie Amazon par exemple, traitait différemment les clients selon leur lieu de résidence et, pour des raisons opaques (l'algorithme n'étant pas accessible), n'offrait pas le service de livraison à la journée pour les habitants des quartiers où la population était majoritairement afro-américaine¹¹⁴. Dans le domaine de la justice,

¹¹⁰ Danielle Keats Citron et Frank Pasquale. 2014. « The Scored Society: Due Process for Automated Predictions ». *Washington L. Rev.* 89(1).

¹¹¹ Michael Aleo et Pablo Svirsky. 2008. « Foreclosure Fallout: The Banking Industry's Attack on Disparate Impact Race Discrimination Claims Under the Fair Housing Act and the Equal Credit Opportunity Act ». *BU Pub. Int. LJ.* 18(1)

¹¹² Cathy O'Neil. 2016. *Weapons of Math Destruction. How Big Data Increases Inequality and Threaten Democracy*. New York: Broadway Book, chap.6.

¹¹³ Idem que la précédente (p.119): « La compagnie a sacrifié un peu d'efficacité pour plus d'équité ».

¹¹⁴ Elizabeth Weis. 2016. « Amazon same-day delivery less likely in black areas, report says ». Dans USA Today. En ligne. <https://www.usatoday.com/story/tech/news/2016/04/22/amazon-same-day-delivery-less-likely-black-areas-report-says/83345684/>

les algorithmes sont de plus en plus utilisés pour prédire les risques de récidive. L'intérêt pour la prédiction des crimes vient du fait que tant la population carcérale que le coût de l'emprisonnement ont fortement augmenté ; une meilleure prédiction des risques de récidive permet de libérer des condamnés qui ont un taux faible de récidive, ou, pour le dire autrement, elle permet de libérer des places dans les prisons. En 2016, l'enquête du site ProPublica a montré que l'algorithme COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) de la Northpointe, inc., utilisé par le système judiciaire en Floride, prédit que les risques de récidive sont deux fois plus élevés pour les criminels noirs que les criminels blancs¹¹⁵.

De manière surprenante, on peut dire en prenant des raccourcis de langage que les SIA sont victimes de biais semblables aux biais cognitifs, comme le biais de confirmation : le traitement discriminatoire de certains groupes renforce non seulement l'inégalité, mais entretient les conditions de la violence sociale. En prédisant que les criminels afro-américains ont deux fois plus de chance de récidiver, en augmentant ainsi le taux et la durée d'incarcération pour cette population, les SIA tendent sinon à créer une situation de discrimination grave, du moins à la perpétuer. Et la machine à discriminer s'autoalimente, ne cherchant dans les données que ce qui confirme ses propres prédictions.

On pourrait objecter que le problème ne vient pas des SIA, que les discriminations ont toujours existé et les algorithmes sont des outils « neutres » pour des politiques qui elles ne le sont pas. Cette objection n'est pas illégitime, elle rappelle qu'il faut distinguer l'outil (les SIA) de son usage (une politique discriminatoire). Il faut cependant faire un examen critique de l'outil lui-même et de ses applications concrètes. Tout d'abord, lorsqu'ils sont développés pour certaines politiques comme l'évaluation de la récidive, les outils sont porteurs des discriminations dénoncées plus haut et ne peuvent être considérés comme « neutres ». Ensuite, les algorithmes ne sont pas infaillibles et leur fiabilité

est très relative selon le domaine envisagé et selon le modèle mathématique utilisé¹¹⁶. Comme le notent les journalistes de l'enquête de ProPublica du 23 mai 2016, si l'algorithme COMPAS donne pour l'ensemble des crimes des résultats plus fiables que le hasard, il donne des résultats erronés pour les crimes violents (ceux qui entraînent pourtant des peines plus lourdes). On pourrait se satisfaire du fait que dans l'ensemble l'algorithme COMPAS soit plus fiable que le hasard, mais dans une démocratie qui reconnaît à chaque personne le droit d'être équitablement traité, ce fait n'est pas pertinent : si dans l'ensemble l'algorithme est fiable, il sacrifie les intérêts fondamentaux de trop nombreuses personnes pour que son usage soit légitime.

Ajoutons enfin que la mise en place de SIA réduit les possibilités de recours, car ils sont considérés, à tort, comme très fiables et non biaisés. Le récit personnel de Virginia Eubanks est édifiant : elle-même confrontée à la décision prise, selon toute vraisemblance par un algorithme, de la suspendre de son assurance médicale, elle a eu la chance de pouvoir compter sur ses connaissances du fonctionnement des algorithmes, sur son employeur et sur ses ressources matérielles.

Les cas que nous venons d'étudier se sont tous produits aux États-Unis. Mais l'État canadien doit se préoccuper des conséquences prévisibles de l'usage des SIA par les administrations publiques au Canada et tirer les leçons des expériences malheureuses dans les autres pays. Si l'automatisation présente un intérêt majeur pour le traitement de millions de dossiers que les administrations traditionnelles peuvent difficilement prendre en charge, les risques de violation des droits fondamentaux des citoyens sont parfois trop importants. Le cas du traitement des dossiers d'immigration est un enjeu stratégique pour l'État canadien. Des centaines de milliers de personnes entrent au Canada chaque année pour des raisons très diverses et cherchent à obtenir un statut de résident temporaire ou permanent. La recherche, menée par le *Citizen Lab* de l'Université de Toronto, souligne les impacts de la prise de décision automatisée pour les demandes

¹¹⁵ Julia Angwin, Jeff Larson, Surya Mattu et Lauren Kirchner. 2016. « Machine Bias ». Dans *ProPublica*. En ligne. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

¹¹⁶ Kate Crawford et Ryan Calo. 2016. « There is a blind spot in AI research ». *Nature* 538 (7625).

d'immigration et la manière dont les erreurs et hypothèses de la technologie pourraient entraîner des conséquences graves pour les immigrants et les réfugiés¹¹⁷. La complexité de nombreuses demandes d'immigration, dans le cas de réfugiés politiques par exemple, pourrait être inadéquatement traitée par les SIA, conduisant à sérieuses violations des droits humains protégés par les différentes conventions internationales dont le Canada est signataire. Les principes éthiques de la Déclaration, le droit québécois et canadien, et le droit international incitent à prendre des mesures de précaution avec les SIA qui sont susceptibles d'engendrer des discriminations sérieuses.

L'identité biaisée : internet et les SIA

Les SIA qui sont utilisés par la grande majorité de la population sont indissociables du fonctionnement plus général de l'internet : ce sont les algorithmes de classement et de recommandation (Google, Amazon, Spotify ou Netflix) ainsi que les réseaux sociaux (Facebook et Twitter par exemple). Dans tous les cas, des algorithmes apprennent des traces que les internautes laissent sur le web et qui signalent leurs comportements habituels, leurs préférences et leurs goûts, leurs idées politiques et leurs conceptions du monde. D'un côté, leurs recherches sur le web et leurs interventions sur les réseaux sociaux, qu'elles soient verbales ou non verbales (mettre en ligne des photos), disent quelque chose de leur « moi », de leur identité, et d'un autre côté, les internautes construisent des représentations de leur identité en fonction des publics auxquels ils s'adressent¹¹⁸. Et ces représentations sont des objets de consommation pour les publics des réseaux sociaux, mais plus largement et plus authentiquement

pour les algorithmes des compagnies en ligne qui récupèrent les données pour vendre des produits, des biens et des services, que ce soit aux individus ou à d'autres compagnies : les données elles-mêmes ou de l'espace pour des publicités ciblées¹¹⁹. Or les algorithmes constituent d'autres intermédiaires, des agents autonomes qui façonnent les représentations et les identités des utilisateurs.

En accord avec les études académiques sur le fonctionnement des algorithmes de classement et les réseaux sociaux, les participants du processus de coconstruction de la Déclaration ont soulevé l'enjeu de l'influence des SIA sur la diversité culturelle et sur les identités qui tendent à la fois à se segmenter selon les groupes et à s'homogénéiser au sein de chaque groupe. Pour mieux comprendre cela, il faut changer de perspective sur les algorithmes et les définir, comme le font Lessig (2006)¹²⁰, Napoli (2014)¹²¹ ou Ananny (2016)¹²², comme des institutions qui gouvernent : « Code is Law », dit Lawrence Lessig, professeur de droit à Harvard et pionnier du mouvement des communs (commons). Autrement dit, les programmes informatiques constituent des lois. En effet, les algorithmes ont le pouvoir de structurer les comportements, d'influencer les préférences, de guider la consommation et de produire du contenu consommable pour des internautes préparés, voire conditionnés. Ce pouvoir s'exerce donc sur l'identité même des internautes et des utilisateurs d'objets connectés, et biaise cette identité en la façonnant.

En classant les contenus et en faisant des recommandations, les algorithmes ont plus fondamentalement la capacité de « structurer les possibilités » offertes aux utilisateurs¹²³ et de créer un univers numérique où les parcours de recherche et d'information sont balisés. Le classement et le filtrage d'une information devenue surabondante

¹¹⁷ The Citizen Lab - University of Toronto. 2018. « Bots at the Gate ». En ligne. <https://ihrp.law.utoronto.ca/sites/default/files/media/IHRP-Automated-Systems-Report-Web.pdf>

¹¹⁸ Lee Humphreys. 2018. « *The Qualified Self: Social Media and the Accounting of Everyday Life* ». Cambridge: The MIT Press.

¹¹⁹ Cathy O'Neil. 2016. *Weapons of Math Destruction. How Big Data Increases Inequality and Threaten Democracy*. New York: Broadway Book, chap.4.

¹²⁰ Lawrence Lessig, 2006. *Code and other laws of cyberspace 2.0*. New York: Basic Books.

¹²¹ Philip M. Napoli. 2014. « Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption ». *Communication Theory* 24(3): 340-360. En particulier, la section « Institutionalité et algorithmes », p. 343 et suivantes.

¹²² Mike Ananny. 2016. « Toward an ethics of algorithms: Convening, observation, probability, and timeliness ». *Science, Technology, & Human Values* 41(1): 93-117.

¹²³ Idem que la précédente (p.97): « Algorithms "govern" because they have the power to structure possibilities ».

auraient pour effet indirect de nuire au pluralisme et à la diversité culturelle : en filtrant les informations, en s'appuyant sur les caractéristiques de leurs profils, les algorithmes augmenteraient la tendance des utilisateurs à fréquenter des personnes et à rechercher les contenus (notamment les opinions et les œuvres culturelles) qui sont a priori conformes à leurs propres goûts et à rejeter l'inconnu¹²⁴. Un individu se retrouve donc enfermé dans une « bulle filtrante », c'est-à-dire dans un espace de recommandations toujours conforme au profil qu'il alimente par son comportement numérique et qui est encouragé par l'environnement numérique qui s'y adapte. Les effets d'une offre culturelle et de contenus, plus abondante que jamais, se voient ainsi paradoxalement neutralisés par un phénomène de réduction de l'exposition effective des individus à la diversité culturelle. Et ce, même si l'individu souhaite une telle diversité.

Une objection pourrait ici être faite : ce que rendent possible les algorithmes, c'est une personnalisation des profils d'utilisation qui, en raison de la diversité des personnes, augmente au contraire la diversité de l'offre. Cette objection pourrait être sérieuse si les algorithmes ne privilégiaient pas des contenus populaires et ne canalisait pas les recherches et les recommandations pour mettre ces contenus en avant. Ce phénomène est renforcé sur les réseaux sociaux par le phénomène bien connu de polarisation qui affecte la formation des opinions et des groupes¹²⁵. Le fonctionnement des réseaux sociaux accélère la polarisation de deux manières :

1. **Tout d'abord parce que les applications mettent à la disposition des utilisateurs des outils qui permettent de filtrer les nouvelles en fonction de ses centres d'intérêt et les personnes avec qui on se connecte en fonction de ses affinités. Le fameux hastag# de Twitter est probablement l'outil de filtrage le plus efficace ; Cass Sunstein évoque la « hastag nation » dans #republic (2017)¹²⁶.**

2. **Ensuite, les algorithmes de ces réseaux sociaux apprennent à repérer ce qui importe aux utilisateurs et ne les alimentent plus qu'avec les informations qu'ils sont censés vouloir connaître. En les recoupant avec des données personnelles laissées sur d'autres sites internet, les algorithmes construisent une chambre d'écho puissante dans laquelle les mêmes personnes, en fonction de leurs intérêts apparents, sont mises en relation, se « connectent », échangent des opinions convergentes, renforcent leurs croyances et consolident leurs caractéristiques collectives.**

Par conséquent, même si une grande diversité de groupes, de fils d'information et de profils de recommandation est générée par les algorithmes des réseaux sociaux, cette diversité est en trompe-l'œil : non seulement la composition interne des groupes tend à s'homogénéiser, mais les groupes deviennent relativement imperméables les uns aux autres. Le fonctionnement des SIA sépare ainsi les individus différents et rassemble les individus semblables. L'inclusion de la diversité appelle au contraire à une diversité inclusive : les personnes différentes sont réunies pour échanger et apprendre de leurs différences.

Pour parvenir à cet objectif, il faut minimalement que les représentations des groupes socialement défavorisés et des minorités de pratique (culturelles, religieuses, sexuelles) ne soient pas caricaturales ni stigmatisantes. Cette condition n'est pas atteinte. Les études académiques sont unanimes : les algorithmes de classement et de recommandation ne sont pas neutres et reflètent les biais en cours dans la société. Plus précisément, ils reproduisent les structures sociales de domination et d'exclusion et contribuent à les renforcer. C'est ce que montre très bien Safiya Umoja Noble dans son livre référence, *Algorithms of Oppression* (2018)¹²⁷ en examinant précisément le fonctionnement de l'algorithme *Google Autocomplete*¹²⁸. La couverture de l'ouvrage illustre le problème (voir figure 1).

¹²⁴ CNIL. 2017. « Comment permettre à l'Homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle ». En ligne. https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf

¹²⁵ Voir les nombreux ouvrages de Cass Sunstein à ce sujet, par exemple : Cass R. Sunstein. 2006. *Infotopia*. Oxford : Oxford University Press.

¹²⁶ Cass R. Sunstein. 2018. # *Republic: Divided democracy in the age of social media*. Princeton: Princeton University Press, p. 79.

¹²⁷ Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

¹²⁸ M. Garber. 2013. « How Google's Autocomplete was ... Created / Invented / Born ». *The Atlantic* 23.

Fig. 1. Détail de la couverture du livre de Safiya Umoja Noble, *Algorithms of Oppression*.



La recherche « Pourquoi les femmes noires sont-elles tellement... » génère les recommandations suivantes : « ... en colère » ; « bruyantes » ; « méchantes » ; « attrayantes » ; « paresseuses », etc. Sans une analyse fine, on voit à l'évidence que l'algorithme *Autocomplete* de Google propose des représentations négatives des femmes noires qui les stigmatisent. Les recherches ouvertes du type : « femmes noires » génèrent quant à elles des propositions de liens pornographiques, réduisant les femmes noires à des objets sexuels¹²⁹. Cela a comme effet de renforcer les stéréotypes culturels¹³⁰ et de dissuader les gens de faire des recherches impopulaires¹³¹.

Ce type de recommandation est problématique pour au moins deux raisons : il renvoie aux autres une image dégradée d'un groupe stigmatisé dans la société et contribue à maintenir les conditions symboliques de la domination sur ce groupe, en renforçant les stéréotypes. En outre, il renvoie une image dégradée aux membres du groupe représenté et affecte ainsi les bases du respect

de soi, le sentiment d'estime de soi et la confiance dans leur valeur. Cette soumission ou sujétion aux représentations de soi définies par autrui, est un facteur majeur de domination par les autres. Les exemples des identités biaisées par les algorithmes abondent. Pour conclure par un exemple plus subtil, évoquons le cas du traducteur de Google du turc à l'anglais :

O bir doctor / O bir hemsire.

La même tournure neutre en turc, avec un pronom personnel indifférent au genre, est traduite de deux manières différentes en anglais associant le rôle de docteur au fait d'être un homme et le rôle d'infirmière au fait d'être une femme : « He is a doctor », « She is a nurse. »¹³² Dans ce cas, le problème est la distribution genrée des rôles sociaux, des professions qui par ailleurs, quels que soient leur importance et leur mérite respectifs, renvoient à une structure hiérarchique de domination où l'homme commande et la femme obéit.

4.2

DÉBIAISER LES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE

Si le fonctionnement actuel des SIA n'est pas neutre et contribue à reproduire les structures sociales de marginalisation, de stigmatisation et de domination, on doit se demander comment corriger la situation et réduire les inégalités qu'il engendre. Il faut dire d'emblée que la neutralité des algorithmes n'est pas le problème à régler contrairement à ce que l'état de la littérature sur le sujet laisse entendre. L'idéal n'est pas la neutralité des algorithmes, ou, du moins, le fonctionnement neutre des algorithmes n'est pas une condition suffisante de l'inclusion de la diversité dans une société.

¹²⁹ Safiya Umoja Noble, p. 19. Idem que la précédente.

¹³⁰ Paul Baker et Amanda Potts. 2013. « Why Do White People Have Thin Lips? Google and the Perpetuation of Stereotypes via Auto-complete Search Forms ». *Critical Discourse Studies* 10 (2): 187-204.

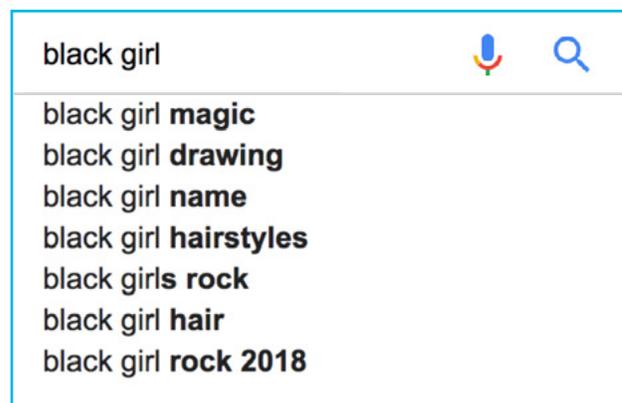
¹³¹ L. Gannes. 2013. « Nearly a Decade Later, the Autocomplete Origin Story: Kevin Gibbs and Google Suggest ». *All Things D*. Accessed January 29: 2014.

¹³² Aylin Caliskan et al. 2017. « Semantics Derived Automatically from Language Corpora Contain Human-Like Biases ». *Science* 356(6334): 183-186; Ryan Calo. 2017. « Artificial Intelligence Policy: A Primer and Roadmap ». *UCDL Rev.* 51: 399.

Quel que soit le sens que l'on donne à la neutralité, elle ne permet pas de corriger ce qui apparaît comme des discriminations non intentionnelles, à moins de prêter des intentions aux SIA ou de démontrer la mauvaise intention des concepteurs et développeurs des algorithmes incriminés. Si on considère qu'un outil est neutre lorsque son usage n'affecte pas l'état de la société, la laisse intacte, on comprend que ce n'est pas ce que l'on recherche pour corriger les discriminations puisqu'on s'efforce au contraire de changer l'état de la société. Si on admet plutôt que la neutralité renvoie à l'usage d'un outil qui ne promeut pas une conception du bien et ne vise pas à créer une situation défavorable pour une partie de la population, on passe encore à côté du problème. En effet, les SIA n'ont pas « l'intention » de reproduire ou de renforcer les discriminations et ne sont pas développés pour cela, mais c'est ce qu'il leur arrive de faire de manière massive en raison de biais dans leur fonctionnement (modèle mathématique ou données d'entraînement).

Il faut donc se départir de cet idéal de neutralité qui n'est pas pertinent à ce niveau de réflexion. Et la raison n'est pas que la neutralité ne peut être atteinte, mais qu'elle n'est pas souhaitable dans la conception des SIA. L'examen critique des SIA a plutôt montré que leur fonctionnement doit être corrigé pour éviter de reproduire des discriminations et de renforcer les conditions de la marginalisation ou de l'exclusion de personnes et de groupes, selon les critères de justice sociale et d'équité qui valent pour les actions humaines. Or ces corrections sont possibles si les humains (programmeurs, explorateurs de données) interviennent. C'est ce que montre Cathy O'Neil avec l'exemple de Xerox puisque l'algorithme de recrutement a été modifié afin de ne pas rejeter les candidatures des personnes habitant des quartiers défavorisés. Il faut donc souligner que la situation s'améliore grâce aux alertes qui sont régulièrement lancées et à l'intervention des êtres humains. Ainsi, la recherche « black women » qui est fournie par Safiya Umoja Noble ne donne plus les mêmes résultats (figure 2).

Fig. 2. Recherche sur le moteur google.com effectuée le 29 octobre 2018.



Mais il reste encore beaucoup de travail à faire comme le montre la figure 3 ci-dessous.

Fig. 3. Recherche sur le moteur google.fr effectuée le 29 octobre 2018.



Comment débiaiser les SIA et rendre leur développement plus inclusif ? La réponse à cette question n'est pas seulement technique, elle est aussi éthique, sociale et politique, et implique que l'on s'intéresse au fonctionnement des SIA.

Un problème avec les données

La première source de biais qui s'impose à l'enquête sur les discriminations est l'élaboration des bases de données sur lesquelles s'exercent les algorithmes. Les données numériques sont comme des ressources naturelles qu'il faut extraire, filtrer, transformer. On parle aujourd'hui de *data mining* (exploration et extraction de données), on compare les données au pétrole. Il existe pourtant une différence fondamentale : à moins de refuser tout réalisme, on doit reconnaître que les ressources naturelles existent même si on ne peut les extraire, et même si on ne peut les percevoir. Les données numériques, en revanche, n'existent pas sans un dispositif de captation et de traitement. Un cœur qui bat n'est pas une donnée ; le rythme cardiaque capté par une montre connectée est une donnée. Et encore, cette donnée n'est pas brute puisque le dispositif de captation (capteur de fréquence cardiaque) doit être couplé à des dispositifs d'interprétation qui produisent une mesure. Les données doivent être générées et interprétées¹³³.

Les algorithmes créent des associations en détectant et en combinant les aspects du monde (caractéristiques, catégories des ensembles de données) qu'ils ont été programmés à voir¹³⁴. Il y a alors deux types de problèmes avec les données : leur qualité et leur extension. La qualité des données peut être affectée négativement par leur étiquetage inadéquat ou moralement inapproprié. Comme ce sont les êtres humains qui doivent eux-mêmes étiqueter la majorité des données d'entraînement, les biais humains comme les présomptions culturelles sont aussi transmis à travers les choix de classifications¹³⁵. Kate Crawford soutient alors qu'il faudrait adopter une démarche qualitative rigoureuse pour examiner et évaluer les sources de données. Même si les méthodologies des sciences sociales peuvent rendre encore plus complexe la compréhension du *big data* (données massives), cela pourrait donner de la profondeur aux données¹³⁶.

Tay, le phénomène GIGO

Tay est un chatbot mis au point par une équipe de développement technologique de Microsoft. Le 23 mars 2016, cette agente conversationnelle a été lancée sur Twitter avec pour fonction d'interagir avec les autres internautes en traitant les messages qu'elle reçoit et en publiant des messages. L'expérience qui devait confirmer que les SIA peuvent désormais passer le test de Turing s'avéra catastrophique. Tay fut « débranchée » moins de 48 h après avoir été lancée.

Le destin de Tay est instructif sur le fonctionnement des algorithmes. En s'éduquant par ses interactions avec d'autres utilisateurs de Twitter, Tay avait très rapidement publié des messages haineux, racistes et sexistes. D'un être humain qui publierait ce genre de message, on dirait qu'il est lui-même raciste et sexiste. Le comportement de Tay s'explique par le fait que les messages qu'elle recevait étaient massivement à caractère raciste et sexiste. En apprenant à partir de données incorrectes (moralement, dans ce cas), l'algorithme Tay donnait des résultats moralement incorrects. Comme le dit l'expression en vogue dans les milieux informatiques : « Garbage in, garbage out » (GIGO).

L'extension des données est l'autre problème auquel il faut faire face. Par là, on entend le fait que les données ne couvrent pas toujours l'ensemble d'un phénomène que l'on veut observer, ou qu'elles sont en trop grande quantité pour une petite partie du phénomène observé. En effet, une des significations

¹³³ Lisa Gitelman (ed.). 2013. « Raw Data is an Oxymoron ». *Cambridge: The MIT Press*.

¹³⁴ Mike Ananny. 2016. « Toward an ethics of algorithms: Convening, observation, probability, and timeliness ». *Science, Technology, & Human Values* 41(1): 93-117

¹³⁵ Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker et Kate Crawford. 2017. « AI NOW Report ». *AI Now Institute at New York University*; Kate Crawford. 2013. « The Hidden Biases of Big Data ». *Harvard Business Review* 1. Voir aussi le rapport du *Big Data Working Group* sous l'administration du Président Obama: Executive Office of the President. 2016. « Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights ».

¹³⁶ Kate Crawford. 2013. « The Hidden Biases of Big Data ». *Harvard Business Review* 1; Adam Hadhazy. 2017. « Biased Bots: Artificial-intelligence Systems Echo Human Prejudices », Dans *Princeton University*. En ligne. <https://www.princeton.edu/news/2017/04/18/biased-bots-artificial-intelligence-systems-echo-human-prejudices>

du biais est statistique et renvoie à l'écart entre un échantillon et une population. Le biais de sélection s'observe lorsque certains membres d'une population ont plus de chance d'être échantillonnés que d'autres. Si cela peut s'expliquer par les biais humains dans la préparation et l'exploration des données, la raison qui est souvent la plus pertinente est que les inégalités systémiques dans la société font en sorte qu'une population est sur-représentée dans les données d'entraînement, et inversement qu'une autre population est sous-représentée¹³⁷. Ainsi, les données sur lesquelles s'entraîne un algorithme peuvent être biaisées ou faussées et présenter un échantillon non représentatif qui a été pauvrement défini avant usage¹³⁸. Un bon exemple est celui des SIA de reconnaissance faciale : plus il y a de visages de personne blanche dans les données d'entraînement, plus le système sera performant pour cette partie de la population¹³⁹. En revanche, dès lors que la population blanche est sur-représentée, d'autres populations, comme les Afro-Américains, sont alors sous-représentées. Le résultat s'avère alors très problématique et a une tendance à la confusion des visages, et même à l'association des visages humains avec des faces de singe, comme dans le très regrettable incident de l'algorithme de Google qui attribuait à des personnes noires l'étiquette (tag) de gorille¹⁴⁰.

Ce phénomène devient dramatique dans le système judiciaire. Aux États-Unis, où l'on utilise déjà différents types de SIA pour faire la prédiction de récidive, le problème principal réside, outre la qualité pauvre des données, dans le manque de données probantes¹⁴¹. En effet, si les crimes d'une partie de la population (disons les Afro-Américains) sont mieux documentés et archivés que les crimes d'une autre partie de la population (disons les Blancs), les premiers seront davantage pénalisés que les seconds, alimentant alors un « cycle de traitements discriminatoires »¹⁴². C'est le problème rencontré

avec un outil de police prédictive comme *PredPol* qui a été conçu selon un modèle mathématique développé pour les risques de séismes, mais qui fonctionne avec un ensemble de données non représentatif.

Faire parler les algorithmes

Si les discriminations s'expliquent en grande partie par une collecte et une extraction défailtantes des données, elles trouvent aussi leur source dans l'algorithme lui-même, son code et son modèle mathématique. Les algorithmes, contrairement à l'ordinateur (l'infrastructure informatique), ne sont pas universels au sens de Turing, c'est-à-dire qu'ils ne réalisent que la tâche pour laquelle ils ont été conçus et ont des objectifs définis par les programmeurs ; un ordinateur est une machine universelle au sens où il est capable de réaliser différentes tâches, mais il a alors besoin de différents algorithmes spécialisés pour y parvenir. C'est pourquoi on peut estimer que les SIA qui engendrent des conséquences discriminatoires sont également en cause. Pour un même ensemble de données, deux algorithmes ayant des paramètres, un modèle mathématique et des objectifs différents généreront des ensembles de résultats différents. On l'a vu avec l'exemple de la compagnie Xerox.

Imaginons que pour éviter une stigmatisation de populations cibles par les algorithmes de classement et de recommandation, on s'entende sur l'objectif suivant : pour une recherche donnée, l'algorithme ne devrait pas fournir toujours les mêmes réponses (dans une période pendant laquelle il n'est pas mis à jour). Par exemple, quand on cherche « femmes noires », on ne devrait pas être dirigé vers des recommandations pornographiques, mais en outre on ne devrait pas non plus se voir proposer toujours les mêmes recommandations

¹³⁷ Alex Campolo et al. 2017. « AI NOW 2017 Report ». *AI Now Institute at New York University*. En ligne. https://ainowinstitute.org/AI_Now_2017_Report.pdf

¹³⁸ Kate Crawford. 2017. *Neural Information Processing Systems (NIPS)*. En ligne. https://www.youtube.com/watch?v=fMym_BKWQzk

¹³⁹ Ryan Calo. 2017. « Artificial Intelligence Policy: A Primer and Roadmap ». *UCDL Rev.* 51: 399.

¹⁴⁰ Alistair Barr. 2015. « Google mistakenly tags black people as 'gorillas,' showing limits of algorithms ». *The New York Times*.

¹⁴¹ Matt Ford. 2015. « The Missing Statistics of Criminal Justice ». Dans *The Atlantic*. En ligne. <http://www.theatlantic.com/politics/archive/2015/05/what-we-dont-know-about-mass-incarceration/394520/>

¹⁴² « AI for the Common Good ». 2018. En ligne. <https://weforum.ent.box.com/v/AI4Good?platform=hootsuite>

« cheveux », « cheveux longs », qui ont remplacé les suggestions dégradantes, mais qui construisent des stéréotypes. On peut alors imaginer l'introduction d'un paramètre de « hasard », un paramètre aléatoire

dans l'algorithme. En procédant ainsi, on résout aussi le problème des bulles filtrantes qui ont un effet sur la diversité et sur l'identité des utilisateurs qui sont enfermés dans un profil d'utilisation.

PARAMÉTRER LA SÉRENDIPITÉ (LA FORTUITÉ)

Le terme de sérendipité vient de l'anglais *serendipity*, un mot qui a été formé par l'écrivain britannique Horace Walpole en 1754¹⁴³. Ce terme renvoie au fait de pouvoir faire une découverte utile par accident, sans l'avoir cherchée. Certaines des plus grandes découvertes scientifiques ont ainsi été faites par accident, comme la pénicilline par Alexander Fleming. Mais la sérendipité, ce n'est pas seulement le hasard ; c'est la possibilité de faire une découverte par hasard et cela doit être facilité par une structure institutionnelle : par exemple, donner du temps aux chercheurs, favoriser les rencontres, ne pas exercer une pression trop importante¹⁴⁴ sur les publications qui absorbent le temps de recherche, etc. De même, les algorithmes de recommandation sont des architectures de choix qui peuvent, ou non, faire place au parcours fortuit et à la découverte.

Nul mieux que l'écrivain Umberto Eco n'a exprimé ce lien entre architecture (de choix) et fortuité. Dans sa conférence sur les bibliothèques, prononcée à Milan en 1981, il déclare :

« Dans une bibliothèque où tout le monde circule et se sert, il y a toujours des livres éparpillés qui n'ont pas été remis sur les rayons [...] Ce type de bibliothèque est à ma mesure, je peux décider d'y passer une journée dans la plus pure joie, je lis les journaux, j'emporte les livres au bar, puis je vais en chercher d'autres, je fais des découvertes. J'étais entré là pour m'occuper, mettons de l'empirisme anglais, et au lieu de cela je me retrouve chez les commentateurs d'Aristote, je me trompe d'étage, je pénètre dans une section où je ne pensais pas entrer, de médecine par exemple, et soudain je tombe sur des ouvrages traitant de Galien, avec des références philosophiques donc. Dans ce sens, la bibliothèque devient une aventure. »

¹⁴³ Pour l'histoire de ce concept, voir Robert K. Merton et Elinor Barber. 2011. *The travels and adventures of serendipity: A study in sociological semantics and the sociology of science*. Princeton (NJ): Princeton University Press.

¹⁴⁴ Umberto Eco. 1986. « De Bibliotheca ». *Μνήμων* 11 337-340.

Si le paramètre est connu et si on peut en mesurer l'impact à partir de tests, on aurait là un algorithme qui évite les bulles filtrantes et les discriminations sans avoir à corriger, après coup et pour des raisons peu évidentes, les résultats de l'algorithme. Prenons l'exemple de la recherche de Safiya Umoja Noble : « Why are black woman so... ». Aujourd'hui Google ne suggère plus la réponse « lazy » (paresseuses). Pourtant, il pourrait être aussi utile de tomber sur cette recommandation qui renverrait à une page où, au lieu de voir une liste de liens vers des publications racistes, on verrait apparaître un lien vers *Le droit à la paresse* de Paul Lafargue (1880). Remettre du hasard, favoriser la sérendipité, quoique cela semble contraire aux buts de la programmation algorithmique, est parfaitement en ligne avec l'objectif de lutter contre les stéréotypes. On trouve d'ailleurs cette idée explicitement formulée par l'inventeur du #hashtag pour Twitter, Chris Messina¹⁴⁵.

Pour s'assurer que les algorithmes ne sont pas biaisés, il faut qu'ils ne soient ni des boîtes noires (*black boxes*), ni des boîtes silencieuses. On parle de boîtes noires pour signaler le fait que le code des algorithmes privés est inaccessible, caché, gardé secret par les compagnies qui les développent. Une des raisons est que l'algorithme constitue une « recette secrète », cruciale à leur affaire et qu'il s'agirait d'une question de propriété intellectuelle¹⁴⁶, ce que l'on peut admettre¹⁴⁷. Mais l'idée de *black box* a une autre connotation : on soupçonne que les compagnies ne veulent tout simplement pas être tenues pour responsables d'algorithmes qui engendrent des discriminations. La façon la plus efficace de se protéger et de maintenir leur modèle d'affaires consiste, pour les entreprises, à déclarer que le fonctionnement des algorithmes est incompréhensible dans le détail et que s'il est arrivé un résultat malheureux, elles ne pouvaient pas le prévoir ni l'empêcher. Présentés comme des boîtes noires, les algorithmes sont protégés de

toutes enquêtes extérieures à l'entreprise qui le développe ou l'utilise. On comprend que cela inspire des craintes et des fantasmes face à la manipulation des entreprises privées¹⁴⁸. Si les individus sont de plus en plus transparents face aux entreprises et gouvernements, la technologie qui facilite cela devient de plus en plus opaque.

Or, si on peut accepter que les compagnies ne veuillent pas divulguer publiquement les codes, on comprend moins bien que les algorithmes ne soient pas accessibles à des autorités compétentes, qu'elles soient publiques ou publiques-privées. Quand les discriminations portent atteinte aux droits fondamentaux des personnes, les pouvoirs publics ont en réalité une obligation d'enquête et de sanction. Par ailleurs, dans le cas des algorithmes publics, un consensus se dégage pour affirmer que leur code devrait être accessible et ouvert.

Ces boîtes noires sont par ailleurs « silencieuses » en ce sens qu'elles n'offrent aux utilisateurs et aux personnes soumises aux procédures algorithmiques aucune information sur le fonctionnement, les objectifs et les paramètres des SIA, ni aucune justification des décisions prises, ou fortement influencées par des SIA. Ce mutisme des SIA, ou des responsables de leur conception et de leur développement, est particulièrement problématique dans une société démocratique qui prône l'inclusion et la justification. C'est en tout cas le sentiment de l'ensemble des participants au processus de coconstruction de la Déclaration, et cela touche une préoccupation qui rejoint la plupart des chercheurs en éthique et en sciences sociales. Une citoyenne propose par exemple qu'on puisse toujours demander une explication compréhensible d'une décision. Les parties prenantes comme l'Ordre des ingénieurs du Québec ont aussi appelé à faciliter la compréhension des décisions algorithmiques.

¹⁴⁵ Cité par Cass R. Sunstein. 2018. *# Republic: Divided democracy in the age of social media*. Princeton: Princeton University Press, p. 79.

¹⁴⁶ Cathy O'Neil. 2016. *Weapons of Math Destruction. How Big Data Increases Inequality and Threaten Democracy*. New York: Broadway Book.

¹⁴⁷ Certains critiquent pourtant la propriété intellectuelle et les normes professionnelles qui font que les algorithmes restent privés, et demandent des codes transparents. Voir Mike Ananny. 2016. « Toward an ethics of algorithms: Convening, observation, probability, and timeliness ». *Science, Technology, & Human Values*. 41(1): 93-117.

¹⁴⁸ Voir à ce sujet Frank Pasquale. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.

Faire parler les algorithmes, cela implique trois choses :

1. **Que les concepteurs de l'algorithme comprennent son fonctionnement (cela paraît trivial, mais cette condition permet de contrecarrer les stratégies de déresponsabilisation des concepteurs) ;**
2. **Que les concepteurs et développeurs soient capables de formuler les paramètres et les objectifs de l'algorithme dans un langage compréhensible pour des personnes lettrées, mais non spécialistes, et qu'ils le fassent ;**
3. **Que les compagnies qui développent ou utilisent un algorithme publient régulièrement des rapports d'impact sociétal (en l'occurrence sur la manière dont il affecte les groupes défavorisés et précaires).**

Comme les algorithmes des SIA sont très complexes et leur comportement difficile à comprendre, même pour les spécialistes¹⁴⁹, les chercheurs s'entendent pour appeler à la mise en place de procédures de test qui permettent d'évaluer les résultats et d'éliminer *ex post* les résultats indésirables. Cela implique aussi que des audits puissent être menés avant la commercialisation d'un algorithme et pendant sa mise en service¹⁵⁰.

Représentativité et inclusivité

Pour assurer une IA inclusive, on ne doit pas seulement s'intéresser à la conception et l'entraînement des algorithmes, mais aussi aux conditions matérielles dans lesquelles ils sont développés. Il faut en particulier examiner les possibles discriminations sociales qui affectent (ou sont produites par) le milieu de la recherche et du développement industriel de l'IA. Il y a deux raisons de s'y intéresser, l'une est instrumentale, l'autre est

déontologique.

La première raison que l'on peut invoquer pour justifier l'objectif de l'inclusion de la diversité dans le milieu du développement de l'IA est que la diversité est une condition propice à l'innovation scientifique et technologique. L'homogénéité du milieu est un facteur de conservatisme scientifique et intellectuel en général. Inutile de développer cet argument ici ; on trouvera chez un auteur comme John Stuart Mill, un plaidoyer défendant les bienfaits épistémiques et moraux de la diversité. C'est aussi l'une des raisons qui expliquent le choix d'un processus ouvert et délibératif pour l'élaboration de la Déclaration de Montréal pour une IA responsable. Mais avant de passer à la raison déontologique, il faut ajouter que l'inclusion de la diversité dans le milieu de l'IA permet aussi de sensibiliser les développeurs des SIA aux enjeux de l'inclusion et des discriminations. En effet, une des explications des biais dans les SIA, que l'on a pour l'instant écartée, est celle des biais des programmeurs eux-mêmes. Utiliser le masculin « programmeur » dans ce cas est tout à fait approprié dans la mesure où l'on constate que la grande majorité des chercheurs et des développeurs en IA sont des hommes. Dans le contexte nord-américain, il faut ajouter que ce sont des hommes blancs, bien payés, avec une éducation technique similaire¹⁵¹. On peut penser que leurs intérêts et leurs expériences de vie influencent leur conception des algorithmes et leur programmation¹⁵². La représentation équilibrée de la diversité des composantes de la société n'est pas une garantie que le développement des algorithmes soit moins biaisé, mais cela paraît néanmoins une condition nécessaire.

Si les raisons instrumentales sont importantes et suffisent à motiver les entreprises, les centres de recherche et les universités à favoriser un développement inclusif de l'IA, la raison déontologique constitue un impératif d'un ordre

¹⁴⁹ Il ne faut pas non plus exagérer la complexité des algorithmes pour leurs concepteurs, ce qui contribue à les percevoir comme des boîtes noires impénétrables, comme le rappelle Taina Bucher. 2018. « *If ... Then. Algorithmic Power and Politics* ». Oxford: Oxford University Press, p. 57.

¹⁵⁰ Voir Cathy O'Neil. 2016. *Weapons of Math Destruction. How Big Data Increases Inequality and Threaten Democracy*. New York: Broadway Book; Campolo, Alex, Madelyn Sanfilippo, Meredith Whittaker et Kate Crawford. 2017. « AI NOW Report ». *AI Now Institute at New York University*; National Science and Technology Council & Office of Science and Technology Policy. 2016. « Preparing for the Future of Artificial Intelligence ». *Office of the President of the President of the United States*.

¹⁵¹ Pour les statistiques dans le contexte des États-Unis, voir le rapport du U.S. Equal Employment Opportunity Commission. 2016. « Diversity in High Tech ».

¹⁵² Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

supérieur. Il s'agit d'une question d'équité sociale. Nous nous intéresserons seulement au cas de la présence féminine dans le milieu de l'IA, par souci de concision, mais l'étude devra comprendre un examen de la situation des minorités ethniques et culturelles. On note que les femmes sont statistiquement moins nombreuses dans les domaines des nouvelles technologies numériques en général et de l'IA en particulier. On pourrait expliquer cette situation par le fait que les femmes sont moins intéressées que les hommes par les sciences informatiques. Évidemment cette réponse ne serait pas suffisante, car il faudrait encore expliquer pourquoi elles seraient moins intéressées que les hommes par le domaine de l'informatique. Or, l'hypothèse la plus crédible est que les femmes sont moins présentes que les hommes dans le domaine de l'informatique, aujourd'hui, non pas en raison d'un manque d'intérêt, ni même d'un manque de formation, mais en raison d'une concurrence forte avec les hommes pour avoir une place dans un secteur social très valorisé et rémunérateur. Cette concurrence est biaisée dès le départ par le fait que les femmes sont découragées de se lancer dans cette compétition.

Il est difficile de corroborer cette hypothèse dans ce chapitre programmatique sur le développement inclusif de l'IA. Cependant, de nombreuses études montrent que les femmes subissent une concurrence faussée qui tourne à l'avantage des hommes. Nous prendrons seulement deux exemples pour conclure ce chapitre. Le premier vient de l'histoire britannique de l'IA remarquablement retracée dans le livre de Marie Hicks au titre éloquent : *Programmed Inequality*¹⁵³. Marie Hicks montre que le Royaume-Uni, au lendemain de la Seconde Guerre mondiale, possédait une classe de travailleurs dans le domaine de l'informatique où la proportion de femmes était très élevée. Les métiers de l'informatique étaient alors mal payés. Mais à partir de 1964, ces métiers avaient été revalorisés et le gouvernement

britannique engagea le pays dans une révolution technologique. Marie Hicks note qu'au même moment l'image des femmes a été utilisée à des fins publicitaires pour vendre des machines, et que progressivement les métiers de l'informatique furent pensés pour les hommes. Le rôle du gestionnaire (manager) devint emblématique de cette révolution technologique et fut associé à l'homme. C'est ainsi que les femmes furent écartées des métiers de l'informatique les plus valorisés.

Le deuxième exemple complètera le premier et illustre le cercle vicieux entre les biais algorithmiques et les discriminations de genre dans le domaine du développement de l'IA. Une étude de la Carnegie Mellon University, menée par Amit Datta, a montré que, sur Google, les femmes avaient moins de chances que les hommes d'être ciblées par des annonces pour des emplois très bien rémunérés (200 000 USD)¹⁵⁴. Comme le remarque Kate Crawford, si les femmes n'ont pas accès à ces annonces, comment pourraient-elles soumettre leur candidature pour ces emplois¹⁵⁵ ? Sachant que les métiers de l'IA sont aujourd'hui très bien rémunérés, le risque est grand que les femmes soient discriminées dès l'annonce de postes à pourvoir. Il est urgent de se préoccuper de cette situation pour que le développement social de l'IA soit véritablement inclusif.

¹⁵³ Marie Hicks. 2017. « Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing ». *The MIT Press*.

¹⁵⁴ Amit Datta, Michael Carl Tschantz et Anupam Datta. 2015. « Automated Experiments on Ad Privacy Settings ». *Proceedings on Privacy Enhancing Technologies* (1): 92–112.

¹⁵⁵ Kate Crawford. 2016. « Artificial Intelligence's White Guy Problem ». *The New York Times* 25.

5. CHANTIER ENVIRONNEMENT : IA et transition écologique, enjeux et défis pour une soutenabilité forte

Plusieurs citoyens ayant participé aux ateliers délibératifs de la Déclaration de Montréal ont rappelé avec force que le développement de l'IA devait se faire de manière soutenable pour la planète. En effet, compte tenu de l'actualité sur l'environnement, avec la crise sur le changement climatique, la transition énergétique, l'épuisement accéléré des ressources naturelles et l'effondrement de la biodiversité, plusieurs enjeux écologiques de la numérisation de la société ont été soulignés, notamment le stockage des données. Certains citoyens ont parlé d'une accumulation outrancière des données et des coûts énergétiques que cela implique, ou de l'accumulation massive et catastrophique des données dans le nuage mondial. Il fut également question de l'enjeu des déchets électriques et électroniques, et de l'obsolescence programmée des objets électroniques de nos vies quotidiennes.

D'autres participants ont également souligné les apports potentiels de l'IA pour la gestion de l'environnement, comme par exemple la surveillance automatique des territoires riches en biodiversité. On a aussi parlé du fait que les applications rendues possibles par l'IA, comme la voiture autonome, ne devaient pas se faire au détriment des expériences de mobilité active (marche, vélo), plus prometteuses pour la transition écologique des villes. Enfin, lors du dernier atelier délibératif d'octobre 2018, une équipe a travaillé directement sur un scénario prospectif de gouvernance algorithmique des comportements individuels et des effets rebonds environnementaux. Ce groupe de discussion a formulé de nombreux enjeux éthiques et démocratiques qui devraient être résolus pour encadrer une telle initiative.

Ces discussions ont ainsi permis de souligner l'importance de l'enjeu environnemental dans le développement mondial de l'IA et ont contribué à enrichir les principes de la Déclaration de Montréal. La pertinence de formuler un nouveau principe sur l'environnement est apparue inéluctable.

PRINCIPE DE DÉVELOPPEMENT SOUTENABLE

Le développement et l'utilisation de SIA doivent se réaliser de manière à assurer une soutenabilité écologique forte de la planète.

Cette exigence d'une soutenabilité forte, revient à souligner que le déploiement des systèmes d'intelligence artificielle (SIA) et ses effets induits sur la société doit être compatible avec les limites environnementales planétaires, le rythme de renouvellement des ressources et des écosystèmes, la stabilité du climat et la non-substituabilité du capital naturel par le capital artificiel¹⁵⁶.

Le *European Group on Ethics in Science and New Technologies*, dans son document *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems* (2018)¹⁵⁷, définit neuf principes éthiques et prérequis démocratiques, dont le neuvième porte sur la soutenabilité. Ce principe se rapproche également d'une logique de soutenabilité forte en recommandant d'assurer « les préconditions de base pour la vie sur la planète », la « préservation d'un bon environnement pour les générations futures », ainsi que « la priorité de la protection de l'environnement ».

Le présent document approfondit ces enjeux environnementaux des SIA. Dans un premier temps, il aborde la question de la contradiction actuelle entre la transition numérique et la transition écologique. Dans un deuxième temps, il précise cet enjeu du point de vue de l'intelligence artificielle en distinguant ce qui relève de l'empreinte écologique de l'IA, avec ses effets environnementaux induits, et de l'IA comme outil au service de la transition écologique. Les recommandations pour une soutenabilité forte des systèmes d'IA en société sont en conclusion de ce rapport sur les chantiers prioritaires.

¹⁵⁴ Pour une présentation de cette notion voir : Dominique Bourg et Augustin Fragnière. 2014. *La pensée écologique. Une anthologie*, Paris, Presses universitaires de France. Chapitre « Enjeux économiques : durabilité faible ou durabilité forte », p. 439-443.

¹⁵⁵ European Commission. 2018. « Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems ». En ligne. https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

5.1

TRANSITION NUMÉRIQUE ET TRANSITION ÉCOLOGIQUE : UNE CONTRADICTION NON RÉSOLUE

Les questions de l'empreinte écologique de l'intelligence artificielle et de l'IA pour la planète (« AI for Earth ») ont récemment été mises à l'agenda des décideurs avec la conférence « AI for Good », alignée sur les objectifs de développement durable des Nations Unies¹⁵⁸ avec le dernier Forum économique mondial (2018), par le lancement du programme "AI for Earth" de Microsoft (2017)¹⁵⁹ et avec le rapport Villani (2018), qui lui consacre un chapitre complet¹⁶⁰.

Cette mise à l'agenda du lien entre intelligence artificielle et environnement est une bonne nouvelle. Elle permet notamment d'approfondir la discussion sur les synergies potentielles et les contradictions entre les deux grandes transitions contemporaines, numérique et écologique¹⁶¹. D'une part, la transition numérique, incluant les mégadonnées, l'intelligence artificielle, l'Internet des objets (IdO) et les nouvelles interfaces, représente actuellement l'une des plus grandes forces de transformation de nos sociétés au XXI^e siècle. D'autre part, la transition écologique est une nécessité incontournable face aux trois enjeux majeurs que sont le changement climatique, l'effondrement de la biodiversité et l'épuisement accéléré des ressources. Ces enjeux sont de plus accompagnés de sérieux problèmes sanitaires et sociaux : fortes inégalités sociales face aux événements climatiques extrêmes, risques pour la sécurité alimentaire dans certaines régions, impacts

sur la santé de la pollution atmosphérique dans les villes (par les activités de combustion qui produisent aussi des GES). Ils posent également un défi de taille : le Jour du dépassement de la Terre, fondé sur le concept de l'empreinte écologique (Rees, 1992), intervient de plus en plus tôt dans l'année. Les derniers rapports du Programme des Nations Unies pour l'environnement (UNEP)¹⁶² et du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC)¹⁶³ indiquent des efforts insuffisants des pays pour réduire leurs émissions de gaz à effet de serre. De plus, l'approche par les limites planétaires (« Planet Boundaries »), qui prend en compte des seuils critiques dont le dépassement pourrait conduire à des changements globaux irréversibles, présente une situation critique. En effet, plusieurs limites sont déjà atteintes et d'autres sont sur le point de l'être¹⁶⁴.

Or, la transition numérique, elle, ne cesse de s'accélérer dans le monde, que ce soit pour les entreprises (ex. industrie 4.0), les villes (smart cities) ou les citoyens (mobilité connectée), avec des disparités de profils de consommation numérique encore fortes. En moyenne en 2018, un Américain possède 10 périphériques numériques connectés et consomme 140 gigaoctets de data par mois, alors qu'un Indien en possède un seul et consomme 2 gigaoctets (The Shift Project, 2018). Les projections sur les acquisitions d'équipements comme les téléphones intelligents ou l'Internet des objets (IdO) par les particuliers et les entreprises indiquent une accélération générale : d'ici 2025, l'association des opérateurs de téléphonie GSMA anticipe une augmentation nette de 3,6 milliards d'utilisateurs de la 4G dans le monde d'ici 2025, et de 1,2 milliard de nouveaux usagers de la 5G¹⁶⁵.

¹⁵⁸ ITU. 2017. « AI for Good Global Summit ». En ligne. <https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx>; ITU. 2018. « AI for Good Global Summit ». En ligne. <https://www.itu.int/en/ITU-T/AI/2018/Pages/default.aspx>

¹⁵⁹ Brad Smith. 2017. « AI for Earth can be a game-changer for our planet ». Dans *Microsoft*. En ligne. <https://blogs.microsoft.com/on-the-issues/2017/12/11/ai-for-earth-can-be-a-game-changer-for-our-planet/>

¹⁶⁰ Cédric Villani. 2018. « Donner un sens à l'intelligence artificielle : Pour une stratégie nationale et européenne ».

¹⁶¹ IDDRI. 2018. « Livre blanc Numérique et Environnement ». En ligne. <https://www.iddri.org/fr/publications-et-evenements/rapport/livre-blanc-numerique-et-environnement>

¹⁶² UNEP. 2017. « Emissions Gap Report ». En ligne. <https://www.unenvironment.org/resources/emissions-gap-report-2017>

¹⁶³ IPCC. 2018. « Special Report on Global Warming of 1.5 °C ». En ligne. <http://www.ipcc.ch/report/sr15/>

¹⁶⁴ Earth Overshoot Day. 2018. En ligne. <https://www.overshootday.org> ; William E. Rees. 1992. « Ecological footprints and appropriated carrying capacity: what urban economics leaves out ». *Environment and Urbanization* 4 (2): 121-130 ; Johan Rockström et al. 2009. « Planetary boundaries: exploring the safe operating space for humanity ». *Ecology and Society* 14 (2): 1-33 ; Will Steffen et al. 2015. « Planetary boundaries: Guiding human development on a changing planet ». *Science* 347(6223) : 1-10.

¹⁶⁵ GSMA. 2017. « Global Mobile Trends ». En ligne. <https://www.gsma.com/globalmobiletrends/>

Cela pourrait offrir des débits allant jusqu'à 10 gigaoctets par seconde (100 fois plus que la 4G) et permettre une intensification de l'usage de la vidéo mobile. En Inde, le taux d'adoption de téléphones intelligents devrait ainsi passer de 45 à 74 % entre 2017 et 2025¹⁶⁶, avec la 4G comme version principale (62 %). Et les objets connectés devraient globalement passer de 9 milliards à 55 milliards dans le monde entre 2017 et 2025. Cela se traduit par une explosion du trafic de données sur le réseau et dans les centres de données. Selon un rapport de Cisco¹⁶⁷, le trafic mondial devrait croître de 25 % par an (de 6,8 zettaoctets en 2016 à 20,6 Zo en 2021), principalement généré par la vidéo (streaming, VOD, cloud gaming) et l'Internet des objets. Le stockage dans les centres de données ne devrait augmenter que de 36 % par an dans le monde (de 286 exaoctets en 2016 à 1,3 Zo en 2021), les données stockées sur des objets connectés seront à 5,9 Zo en 2021, soit 4,5 fois plus importantes que celles stockées dans les centres de données. Le total des données créées (et non nécessairement stockées) atteindra 847 Zo par an en 2021, contre 218 Zo en 2016.

Ko, Mo, Go, To, Po, Eo, Zo... en films HD

Un film en HD occupe autour de 4 Go en mémoire numérique. Les ordinateurs personnels actuels ont souvent un disque dur pouvant stocker 1 To, soit 250 films. Le Zo, qui vaut un milliard de To, est donc équivalent à 250 milliards de films HD. Le total des données créées dans le monde en 2016 était de 218 Zo, soit plus de 7 000 films par habitant de la planète.

Pour communiquer ces données, la technologie 5G, avec un débit de 10 Go/s, permettrait de télécharger l'équivalent de 2 films HD par seconde sur un objet connecté.

Enjeux environnementaux

Les experts du Shift Project¹⁶⁸ soulignent que cette croissance est essentiellement attribuable à des services offerts par quelques grandes entreprises, les GAFAM américains (Google, Apple, Facebook, Amazon et Microsoft) et les BATX chinois (Baidu, Alibaba, Tencent et Xiaomi). Cette croissance se produit à un rythme qui surpasse celui des gains d'efficacité énergétique des équipements, des réseaux et des centres de données. Cette transition est en effet très matérielle, et la réalité des impacts environnementaux, souvent occultés ou méconnus, doit être soulignée.

La production d'un téléphone intelligent engendre de nombreux impacts tout au long de son cycle de vie, de l'extraction des matières — enjeux de biodiversité, de conditions de travail, épuisement de ressources comme les terres rares, qui sont par ailleurs indispensables pour la production d'énergies renouvelables, comme l'indium (utilisé pour les écrans et les cellules photovoltaïques) ou le néodyme (utilisé dans les aimants des générateurs d'éoliennes) — à la fin de vie (problématique des déchets électroniques, dont très peu sont recyclés, en passant par la phase d'utilisation : consommation d'énergie du terminal, mais aussi du réseau et des centres de données). Sur le changement climatique, environ 90 % des impacts d'un téléphone (ex. 32 Kg CO₂eq pour un 5 pouces) ont lieu lors de la phase de production¹⁶⁹. Ceci s'explique par le fait que ces téléphones ont une durée d'utilisation très courte (environ 2 ans), à cause de l'obsolescence programmée. Les impacts de la fabrication apparaissent donc comme très élevés dans la vie d'un appareil. Les processeurs comme les GPU, qui sont utilisés à la fois en jeux vidéo et en intelligence artificielle, sont également

¹⁶⁶ Peter Newman. 2018. « IoT Report : How Internet of Things technology is now reaching mainstream companies and consumers ». Dans *Business Insider*. En ligne. <https://www.businessinsider.com/internet-of-things-report>

¹⁶⁷ Cisco. 2018. « Cisco Global Cloud Index, Forecast and Methodology 2016–2021 ». En ligne. <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf>

¹⁶⁸ The Shift Project. 2018. « Lean ICT. Pour une sobriété numérique ». En ligne. <https://theshiftproject.org/article/pour-une-sobriete-numerique-rapport-shift/>

¹⁶⁹ ADEME. 2018. <https://www.ademe.fr/modelisation-evaluation-impacts-environnementaux-produits-consommation-biens-dequipement>, et The Shift Project (2018). Idem que la précédente.

consommateurs d'énergie¹⁷⁰. Les centres de données consomment également des ressources limitées comme le silicium, de l'énergie et de l'eau pour leur refroidissement. Quant aux objets connectés, ils contribuent à produire des déchets électriques et électroniques, tout en consommant de l'énergie. Des déchets électroniques qui sont en partie réexportés vers des pays en voie de développement où ils sont démontés dans des conditions sanitaires et sociales très difficiles¹⁷¹.

Le rapport Villani (2018)¹⁷² cite un rapport de l'association américaine des industriels du semi-conducteur qui prévoit qu'en 2040, les besoins en espace de stockage au niveau mondial pourraient excéder la production disponible globale de silicium, et que l'énergie requise pour les besoins en calcul devrait également dépasser la production énergétique mondiale¹⁷³.

À plus court terme, les experts du Shift Project indiquent que la part mondiale du numérique dans les émissions de GES est passée de 2,5 % à 3,5 % entre 2013 et 2018, et pourrait atteindre 4 % en 2020 (2,1 GtCO₂eq). Dans un scénario d'accélération non maîtrisée de la transition numérique et de politiques climatiques inchangées, on atteindrait près de 8 % en 2025 (4,1 GtCO₂eq). Ils indiquent également que l'empreinte énergétique du numérique (incluant l'énergie de fabrication et d'utilisation des équipements : serveurs, réseaux, terminaux) augmente actuellement de 9 % par an et capte une part croissante de l'électricité mondiale, ce qui peut compromettre sa décarbonation (abandon des énergies fossiles pour produire des kWh). Ils soulignent enfin la croissance probable de la part du numérique dans la consommation mondiale

d'énergie. Si celle-ci était à 1,3 % en 2013, elle doublait déjà à 2,7 % en 2017. Selon leurs prédictions, elle pourrait être entre 3,2 % et 6 % en 2025, selon le rythme de la transition numérique et les gains d'efficacité énergétique. À 6 %, la part du numérique représenterait une consommation de plus de 25 % de l'électricité mondiale en 2025 !

Le GtCO₂eq : une mesure d'émissions de gaz à effet de serre

Il existe plusieurs gaz à effet de serre. Si le dioxyde de carbone, ou CO₂, est responsable de 76 % du réchauffement planétaire dû à l'activité humaine, d'autres doivent aussi être considérés, comme le méthane CH₄ ou le protoxyde d'azote N₂O¹⁷⁴. Chaque gaz a un potentiel de réchauffement global (PRG) différent. Le CO₂ est considéré comme référence : son PRG est 1. Le méthane, par exemple, a un PRG de 25 : une tonne de CH₄ a donc un impact 25 fois plus grand qu'une tonne de CO₂. Le PRG permet de comparer les émissions de gaz différents, en utilisant la tonne de CO₂ équivalente (tCO₂eq) comme unité.

En 2016, le Canada a produit 704 MtCO₂eq¹⁷⁵, soit l'équivalent de 704 millions de tonnes de CO₂. La même année, le monde produisait autour de 50 GtCO₂eq.

¹⁷⁰ Sur l'efficacité énergétique des processeurs, le "Green 500" (<https://www.top500.org/green500/>) propose un classement mondial des supercalculateurs selon leur efficacité énergétique, en mesurant l'indicateur FLOPS/Watt (puissance de calcul d'une machine en une seconde par énergie consommée). Selon le Green 500, les dix meilleurs processeurs ont une efficacité située entre 11 et 17 GFLOPS/Watt en 2018, contre 2 à 3 GFLOPS/Watt en 2013). Voir sur ce thème: Balaji Subramaniam et al. 2013. Trends in energy-efficient computing: A perspective from the Green500, IEEE. En ligne. <https://ieeexplore.ieee.org/document/6604520>

¹⁷¹ EFFACE. 2015. « Illegal shipment of e-waste from the EU (European Union action to fight environmental crime) ». En ligne. <https://efface.eu/illegal-shipment-e-waste-eu-case-study-illegal-e-waste-export-eu-china>; World Health Organization. 2017. « Children environmental health, electronic waste ». En ligne. <http://www.who.int/ceh/risks/ewaste/en/>

¹⁷² Cédric Villani. 2018. « Donner un sens à l'intelligence artificielle : Pour une stratégie nationale et européenne ».

¹⁷³ SIA. 2015. « Rebooting the IT Revolution, a Call to Action ». En ligne. <https://eps.ieee.org/images/files/Roadmap/Rebooting-the-Revolution-SIA-SRC-09-2015.pdf>

¹⁷⁴ Cf.: EPA. « Global Greenhouse Gas Emissions Data ». En ligne. <https://www.epa.gov/ghgemissions/global-greenhouse-gas-emissions-data>

¹⁷⁵ Cf. : Gouvernement du Canada. 2018. « Greenhouse Gas Emissions ». En ligne. <https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/greenhouse-gas-emissions.html>

Effets rebonds et objectifs de réduction des GES : le nœud de la contradiction

En dynamique, cette tendance générale peut être expliquée par de nombreux effets rebonds¹⁷⁶. Alors que l'efficacité énergétique des équipements s'améliore, plutôt que de verrouiller ces acquis, on consomme proportionnellement plus d'équipements et de services : les données stockées augmentent et les appareils utilisés se diversifient (ex. Internet des objets), la taille des écrans augmente, les applications possibles se renouvèlent sans cesse et le nombre d'équipement des usagers augmente. De plus, ces équipements sont renouvelés à un rythme très rapide, selon des logiques d'obsolescence multiples (logicielle, algorithmique, stylistique, de puissance, programmée). Il en résulte une augmentation des émissions de GES du secteur, des déchets électriques et électroniques croissants, et une pression sur les ressources rares et la biodiversité, notamment lors de l'extraction des matières premières. Avec ces effets rebonds, il n'y a donc pas de découplage entre le développement du numérique d'une part, et sa matérialité et son empreinte écologique d'autre part.

Ces tendances sont en forte contradiction avec les objectifs de réduction des GES adoptés lors de l'Accord de Paris de 2015 pour maintenir le réchauffement de la planète en dessous de 1,5 ou 2 degrés par rapport à l'ère préindustrielle. Cette contradiction s'accroît dans les récentes publications de l'UNEP¹⁷⁷ et du GIEC¹⁷⁸, qui indiquent qu'un effort sans précédent de réduction de notre consommation énergétique et de nos émissions de GES devrait être réalisé dès la prochaine décennie à l'échelle mondiale. Ces rapports démontrent que les émissions mondiales annuelles de GES, actuellement légèrement supérieures à 50 GtCO₂eq par an, devraient être réduites de 10 GtCO₂eq d'ici 2030 pour avoir une chance de tenir l'objectif de 2 °C, et de 20 GtCO₂eq d'ici 2030 pour l'objectif de 1.5°C ! Et dans cette trajectoire à inventer, qui va au-delà des politiques actuelles et des engagements pris par les pays, chaque Gigatonne de CO₂eq émise annuellement compte.

Le numérique au service de la transition écologique

Parallèlement à la problématique de l'empreinte environnementale du numérique, s'ouvre une autre perspective beaucoup plus convergente, par laquelle les applications du numérique opèrent comme des accélérateurs de la transition écologique (Iddri et al., 2018). En plus des réseaux énergétiques intelligents, de la ville ou de l'agriculture intelligente, de nombreuses initiatives innovantes ont trouvé dans le numérique un outil de participation, d'organisation et de partage de connaissances au service de la transition écologique : site web sur les gestes durables ou sur la biodiversité, site web sur les circuits courts alimentaires ou sur le covoiturage, site web de cofinancement des énergies vertes, ou pour sensibiliser à l'obsolescence programmée, ou encore, télétravail et visioconférence.

Ainsi, « Green IT » et « IT for Green » offrent deux voies complémentaires pour penser la convergence et les contradictions entre les transitions écologiques et numériques. C'est cette double approche que nous allons suivre pour aborder les relations entre intelligence artificielle et environnement.

5.2

INTELLIGENCE ARTIFICIELLE ET ENVIRONNEMENT : DÉFIS ET OPPORTUNITÉS

Quels sont les effets spécifiques de l'essor récent des systèmes d'intelligence artificielle (SIA), avec sa forme la plus récente, l'apprentissage machine, sur la transition numérique et pour l'environnement ? Nous analyserons ces effets en adoptant deux perspectives : d'une part, la contribution directe et indirecte des SIA à l'empreinte écologique de la transition numérique, et d'autre part, l'arrivée de nouveaux outils d'inférence prédictive, au service de la transition énergétique et écologique.

¹⁷⁶ Ray Galvin. 2015. « The ICT/electronics question: Structural change and the rebound effect ». *Ecological Economics* 120: 23-31.

¹⁷⁷ UNEP. 2017. « Emissions Gap Report 2017 ». En ligne. <https://www.unenvironment.org/resources/emissions-gap-report-2017>

¹⁷⁸ IPCC. 2018. « *Special Report on Global Warming of 1.5 °C* ». En ligne. <http://www.ipcc.ch/report/sr15/>

5.2.1 EMPREINTE ENVIRONNEMENTALE DIRECTE ET INDIRECTE DES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE (SIA)

Le développement et le stockage de bases de données, le recours à des capteurs, la mise au point des algorithmes en apprentissage machine, l'utilisation de nouveaux processeurs, le développement des robots équipés d'IA, sont tous des exemples de SIA. Ces systèmes représentent une partie des activités et technologies du secteur numérique, qui inclut également les terminaux comme les téléphones, tablettes, ordinateurs, téléviseurs, les activités culturelles comme le visionnage de vidéos, les jeux vidéo, le livre numérique, l'internet, les réseaux et centres de données associés. Du point de vue des impacts directs de leurs activités (consommation d'énergie, émissions de GES, utilisation de ressources, déchets et biodiversité sur le cycle de vie), les SIA représentent ainsi une part des impacts environnementaux du numérique. Plusieurs de ces points ont été soulignés par les participants dans les tables de délibération de la coconstruction de la Déclaration de Montréal IA responsable entre février et octobre 2018.

Toutefois, c'est du point de vue de ses effets indirects sur le secteur numérique mondial que les SIA vont avoir un effet important sur l'environnement. En effet, si on aborde les SIA et ses algorithmes comme des catalyseurs et des accélérateurs de la numérisation de la société, avec des effets rebonds multiples, ces systèmes pourraient avoir un effet critique sur l'environnement. Ce « facteur IA » dans la numérisation de la société a lieu de manières multiples (voir encadré ci-dessous).

Effet catalyseur et accélérateur de l'IA sur la numérisation de la société

INTENSIFICATION DES USAGES ACTUELS :

que ce soit la capture de l'attention par des recommandations personnalisées, la génération de nouvelles images et vidéos par GANs (« Generative adversarial networks »), la réalité augmentée et virtuelle, les promesses de gains de productivité de l'industrie 4.0 ou la promesse d'une ville plus intelligente, l'IA rend le numérique plus « désirable » et intensifie ses usages actuels.

EXPANSION NUMÉRIQUE SUR DE NOUVEAUX OBJETS ET SERVICES :

services prédictifs et assistants personnalisés connectés, objets domestiques connectés avec interaction vocale, cobots (robots collaborateurs), voitures autonomes avec capteurs vidéo ; l'IA permet au numérique de renouveler l'identité des objets et services, tout en conduisant à une explosion des données générées, transmises et stockées.

EFFETS ENVIRONNEMENTAUX INDUITS SUR D'AUTRES PRATIQUES :

les recommandations personnalisées par IA des plateformes collaboratives (ex. échange de maisons, achats de produits d'occasion, e-commerce) peuvent provoquer des effets environnementaux induits : plus de transport, une obsolescence accélérée des produits, etc.

ACCÉLÉRATION DU RYTHME DE RENOUVELLEMENT DES ÉQUIPEMENTS

pour avoir **PLUS DE PUISSANCE** et pour pouvoir utiliser les dernières applications en intelligence artificielle. La course à la 5G pour les téléphones intelligents va dans ce sens, ce qui conduira à une pression encore plus forte sur les ressources et l'environnement.

Par cet effet structurant de valorisation, d'intensification et d'expansion des activités numériques actuelles, et d'accélération du rythme de renouvellement des équipements, on peut donc anticiper que les SIA généreront des impacts environnementaux beaucoup plus importants que ceux du numérique actuel en intensifiant et en amplifiant les effets rebonds déjà constatés dans la partie précédente.

Soutenabilité forte

Face à cette évolution, des recommandations sont formulées dans ce document pour que les SIA et leurs effets environnementaux directs et indirects répondent à l'exigence d'une soutenabilité forte, compatible avec les limites environnementales planétaires, le rythme de renouvellement des ressources et des écosystèmes, la stabilité du climat et la non-substituabilité du capital naturel par le capital artificiel¹⁷⁹.

Trois grandes pistes pour une soutenabilité forte des SIA

Ces trois pistes sont les suivantes : la première regroupe des initiatives d'information et de littératie environnementale sur le numérique, pour permettre aux citoyens et acteurs institutionnels d'avoir une plus forte autonomie et une meilleure capacité d'initiative. La deuxième consiste en des démarches d'écoconception pour les entreprises qui développent des SIA. La troisième réunit différentes politiques publiques structurantes pour une soutenabilité forte des SIA. Nous exposons ici leur logique et présentons des exemples inspirants. Ces pistes seront synthétisées dans une liste de recommandations dans la troisième partie de ce document.

I/ SYSTÈMES D'INFORMATION : BIEN INFORMER, MAIS AUSSI CONSEILLER

Des dispositifs d'information sur l'empreinte écologique des produits existent avec des écolabels de type 1 (ISO 14024), qui garantissent au consommateur une information sur la performance environnementale du produit sur son cycle de vie : l'Écologo canadien, l'écolabel européen et d'autres écolabels, dits de type 3 (ISO 14025), plus utilisés dans les relations entre clients et fournisseurs, présentent un résumé d'analyse de cycle de vie (ACV) du produit : c'est le cas du dispositif EPD (Environmental Product Declaration) qui présente une ACV vérifiée par un tiers. D'autres labels écologiques sont utilisés dans le secteur des produits informatiques : le standard IEEE1680 et l'EPEAT. D'autres enfin visent spécifiquement la phase d'utilisation d'appareils électroménagers, gros consommateurs d'énergie (réfrigérateurs, lave-linge, etc.) : le label Energy Star ou l'étiquette énergie obligatoire sur le marché des produits électroménagers européens qui positionne l'efficacité énergétique d'un appareil sur une échelle de performance en 7 à 10 classes.

Pour les systèmes d'IA, qui combinent des bases de données, des capteurs, des interfaces, des produits et des services dans une solution intégrée, et qui peuvent avoir des effets indirects sur le cycle de vie (ex. un centre de données utilisant du kWh produit par énergie fossile), ainsi que des effets induits sur la numérisation de la société, des écolabels spécifiques prenant en compte l'ensemble du cycle de vie devront être élaborés. Face au problème de l'obsolescence programmée des appareils qui crée une pression sans précédent sur les ressources et la biodiversité, ces écolabels devront notamment inclure des critères liés à la prolongation de la durée de vie des appareils utilisés par l'ensemble du système d'activités mobilisées par le SIA (ex. sur les manières écologiques de mettre les capteurs de données comme les interfaces utilisateurs à niveau sans devoir les jeter). Face aux risques d'impacts liés au traitement des

¹⁷⁹ Pour une présentation de cette notion voir : Dominique Bourg et Augustin Fragnière. 2014. *La pensée écologique. Une anthologie*, Paris, Presses universitaires de France. Chapitre « Enjeux économiques : durabilité faible ou durabilité forte », p. 439-443.

données massives, une attention particulière sur l'infrastructure de collecte et de stockage des données devra faire partie du diagnostic de cycle de vie. Un label « SIA écologique et social » pour les entreprises développant des systèmes d'intelligence artificielle devrait ainsi être développé et utilisé comme un critère de sélection lors des appels d'offres publics et privés, et dans les relations avec les consommateurs.

De plus, informer sur la qualité écologique d'un SIA ne suffit pas. Une pédagogie active sur les usages écologiques des SIA et une littératie environnementale sur les SIA devraient être déployées auprès des citoyens comme des entreprises et des administrations publiques : sur l'obsolescence programmée, sur la capture de l'attention et sur les effets rebonds. Par exemple, Iddri et al. (2018)¹⁸⁰ souligne que les véhicules autonomes de demain, qui utiliseront des SIA, pourraient être partagés dans une logique de transport en commun. Mais ils pourraient aussi rester la propriété individuelle de personnes qui profiteront d'un confort accru pour habiter toujours plus loin de leur lieu de travail et tourner le dos aux transports collectifs. Autre exemple, les recommandations personnalisées par algorithmes prédictifs sur les sites internet culturels tentent de capturer l'attention des usagers : une facilité de déconnexion devrait toujours être offerte tout comme un apprentissage à la déconnexion et à l'autonomie devrait être offert à chaque citoyen. La manière d'utiliser les SIA sera donc déterminante pour juger de leur impact écologique.

Les livrets d'information pour le grand public de l'ADEME sur les enjeux environnementaux du numérique donne un exemple intéressant de ce type d'initiative de sensibilisation¹⁸¹. Les lieux de déploiement de cette sensibilisation doivent aussi

être choisis : dans les écoles, les bibliothèques publiques, les commerces, sur les sites internet vendant ou utilisant des SIA, entre autres.

Enfin une grande base d'information de référence, publique, gratuite et accessible sur les impacts environnementaux sur le cycle de vie des SIA et du numérique devrait être mise en place aux échelles locale, nationale et internationale. L'initiative de « The Shift Project » sur un Répertoire environnemental du numérique ou les publications publiques de l'ADEME sur les impacts environnementaux des biens de consommation et d'équipements¹⁸², vont par exemple dans ce sens.

II/ ÉCOCONCEPTION : UNE APPROCHE CONSÉQUENTIELLE POUR LES SIA ?

Depuis plus de vingt ans, des démarches d'écoconception, qui permettent d'intégrer les critères écologiques et sociaux dès la phase de conception et de développement des produits et services¹⁸³, se sont diffusées dans de nombreux secteurs. Dans le secteur du numérique, des initiatives et référentiels d'écoconception prenant en compte le cycle de vie physique ont également vu le jour : les *Principles for Digital Development* comportent un chapitre « Build for sustainability »¹⁸⁴, et un ouvrage a été publié sur l'écoconception des sites internet¹⁸⁵.

Compte tenu des enjeux environnementaux directs et indirects associés aux SIA, un référentiel d'écoconception des SIA qui permet aux entreprises développant des solutions d'intelligence artificielle (ex. un algorithme de recommandation, un outil d'aide à la décision, un robot domestique, un système pour la ville intelligente) serait très pertinent. Le sous-comité ISO/IEC JTC 1/SC 42

¹⁸⁰ IDDRI. 2018. « Livre blanc Numérique et Environnement ». En ligne. <https://www.iddri.org/fr/publications-et-evenements/rapport/livre-blanc-numerique-et-environnement>

¹⁸¹ ADEME. 2017. « La face cachée du numérique ». En ligne. <https://www.ademe.fr/face-cachee-numerique>

¹⁸² The Shift Project. 2018. « Lean ICT. Pour une sobriété numérique ». En ligne. <https://theshiftproject.org/article/pour-une-sobriete-numerique-rapport-shift/> et ADEME (2018), Idem que la précédente.

¹⁸³ Voir par exemple la norme ISO 14006 2011. « Systèmes de management environnemental - Lignes directrices pour intégrer l'écoconception » dans Carlo Vezzoli et Ezio Manzini. 2018. « Design for Environmental Sustainability ». London: Springer, p.4

¹⁸⁴ Principles for Digital Development. En ligne. <https://digitalprinciples.org/principle/build-for-sustainability/>

¹⁸⁵ Frédéric Bordage. 2015. Eco-conception web / les 115 bonnes pratiques. Paris : Editions Eyrolles.

¹⁸⁶ ISO. 2017. « ISO/IEC JTC 1/SC 42 : Artificial Intelligence ». En ligne. <https://www.iso.org/committee/6794475.html>

récemment créé à l'ISO¹⁸⁶ pour développer un cadre normatif international sur l'intelligence artificielle et son écosystème pourrait aussi se saisir de cette question de l'écoconception des SIA, comme des autres enjeux éthiques de l'IA, en coordination avec le comité technique ISO/TC 207 qui travaille sur les normes de management environnemental ISO 14000.

Quels seraient les enjeux spécifiques de l'écoconception des SIA ? Comment intégrer des critères environnementaux dans l'apprentissage machine et les applications qui en résulteraient ? Ce type de travail devra être développé par des comités multidisciplinaires et multipartites. On se contentera de souligner ici quelques pistes. La première est d'adopter une approche pour prendre en compte les impacts du cycle de vie sur tout l'écosystème. Cette approche permettrait de développer et de faire fonctionner un système d'IA sans provoquer des transferts d'impact, comme l'utilisation des équipements pour la collecte des données, le fonctionnement des centres de données, l'utilisation d'énergies renouvelables aux étapes les plus intenses en énergie sans détourner des ressources prioritaires pour la transition écologique, et l'extraction des matières premières et la fin de vie des équipements. La deuxième serait de réaliser un examen critique du service rendu par le SIA et de ses effets induits pour éviter de provoquer des effets rebonds environnementaux (ex. éviter une capture de l'attention qui pose des enjeux d'autonomie des usagers, mais aussi de surconsommation énergétique). Une autre piste serait de générer une démarche d'analyse de cycle de vie conséquente pour estimer les effets environnementaux induits par l'adoption des SIA sur la société.

Ces démarches d'écoconception pourraient être stimulées par des démarches d'audits environnementaux. L'institut AI Now¹⁸⁷ a mis en avant l'importance d'audits éthiques pour les SIA dans les secteurs les plus sensibles (éducation, justice, santé), en s'inspirant notamment du droit de l'environnement. Plus que d'opérer seulement un parallèle avec le secteur environnemental, le secteur de l'IA pourrait aussi réaliser des audits

sur les pratiques d'écoconception des SIA. C'est une piste également formulée par l'organisme *Data and society* dans un document de réflexion¹⁸⁸. Des plateformes d'évaluation environnementale des SIA, comme <http://www.ecoindex.fr> sur l'empreinte environnementale des sites internet, pourraient aussi représenter une piste intéressante.

Pour soutenir ces démarches d'écoconception, des programmes de formation et des ressources devraient être déployés : accès gratuit à des données environnementales de cycle de vie de qualité, bases de données environnementales publiques pour permettre aux acteurs du numérique d'analyser leurs impacts environnementaux, réseaux de partage de bonnes pratiques et MOOC (« Massive Open Online Course » ou formation en ligne ouverte à tous) sur l'écoconception des SIA.

III/ POLITIQUES PUBLIQUES ET POLITIQUES DE RECHERCHE : QUEL « GIEC » POUR L'IA ?

Des politiques publiques d'achats verts et responsables, pour intégrer systématiquement des clauses écologiques et éthiques dans les appels d'offres publics sur les SIA, devraient être élaborées. Par exemple, sur la prolongation de la durée de vie des équipements, l'interdiction de l'obsolescence programmée (effective dans un pays comme la France, avec la *Loi sur la transition énergétique* de 2015) et la promotion des principes d'économie circulaire, pour verdir la chaîne de valeur de l'IA. Des principes comme l'écoconception des centres de données devraient également être systématiquement promus par les pouvoirs publics.

Ensuite, une grande politique de recherche interdisciplinaire sur les liens entre IA, numérisation et transition écologique devrait être organisée aux niveaux national et international. Le rapport Villani (2018) préconise dans le même sens de « mettre en place un lieu dédié à la rencontre de la transition écologique et de l'IA ». Ce travail pourrait être organisé dans un sous-groupe dédié du GIEC actuel (Groupe d'experts intergouvernemental sur

¹⁸⁷ Dillon Reisman et al. 2018. « Algorithmic Impact Assessments : A practical framework for public agency accountability ». En ligne. <https://ainowinstitute.org/aiareport2018.pdf>

¹⁸⁸ Data & Society, 2018. « (Closed) Call for Applications: Environmental Impact of Data-Driven Technologies Workshop ». En ligne. <https://datasociety.net/blog/2018/07/03/call-for-applications-environmental-impact-of-data-driven-technologies-workshop/>

le changement climatique), dans son volet Mitigation, ou de ce qui serait un nouveau « GIEC » sur l'éthique de l'IA. Cette politique de recherche devrait aborder des chantiers aussi variés et importants que l'impact environnemental des centres de données (et leur localisation dans le monde pour éviter de détourner des ressources locales), la gestion prévisionnelle des métaux rares pour la transition écologique, les déchets électriques et électroniques de l'Internet des objets et l'économie circulaire, le contrôle des effets rebonds et de l'obsolescence technologique, logicielle et algorithmique accélérée, les bénéfices environnementaux et les enjeux éthiques du stockage ADN, l'apprentissage machine à très faible consommation d'énergie, ou encore l'enjeu émergeant du smog électromagnétique et de la santé environnementale avec l'arrivée de la 5G dans les villes.

5.2.2 DE NOUVEAUX OUTILS PRÉDICTIONNELS POUR LA TRANSITION ÉCOLOGIQUE

Les technologies numériques sans IA offrent déjà de nombreux outils au service de l'environnement, par exemple un site web de partage de connaissances sur l'écologie, un site sur les circuits courts alimentaires, la possibilité de faire du télétravail ou de participer à une réunion sans devoir se déplacer grâce à la visioconférence, ou encore, une plateforme de vélos en libre-service ou d'autopartage. Dans la même logique, des SIA aussi offrent une nouvelle gamme d'outils face à la crise écologique. Des offres de solutions labellisées « IA pour la planète » (AI for Earth) ont récemment vu le jour. Celles-ci s'appuient sur les propriétés spécifiques de l'IA, comme proposer des inférences prédictives en apprentissage supervisé, ou réaliser des catégorisations de données massives en apprentissage non supervisé. Ces propriétés permettent de développer des outils au service de l'environnement :

1. un nouvel outil de connaissance prédictive sur les enjeux environnementaux et sociaux (ex. sur la biodiversité, le changement climatique, la productivité agricole, les événements climatiques extrêmes, les migrations);

2. un nouvel outil d'optimisation prédictive (ex. pour les transports urbains, l'énergie dans les bâtiments, les *smart grids* énergétiques, l'agriculture);
3. un nouvel outil pour réguler de façon prédictive les effets environnementaux des acteurs économiques, en particulier ceux qui relèvent de l'effet rebond.

Quatre grandes pistes sur les SIA pour la transition écologique

I/ L'IA COMME OUTIL DE CONNAISSANCE AU SERVICE DE LA TRANSITION ÉCOLOGIQUE

Le traitement de données massives par IA peut permettre de mieux modéliser et comprendre l'écosystème terrestre. Le rapport Villani (2018, page 127, op. cit.) présente ainsi deux projets illustratifs de ce type de contribution de l'IA à l'environnement. Le projet « Tara Oceans », qui permet de collecter et ouvrir des données massives sur l'océan pour comprendre et modéliser un biome planétaire (la biodiversité et des services écosystémiques de l'océan). De même la recherche sur le climat et la météo, pour une meilleure prévention climatique et des risques climatiques (ex. pour les zones habitées, les écosystèmes, l'agriculture).

Par exemple, l'agriculture durable ou biologique peut être très sensible aux événements climatiques extrêmes et au réchauffement (nouveaux ravageurs) qui peuvent entraîner des pertes de récoltes et altérer la sécurité alimentaire d'une région. Si l'IA peut contribuer à une prévision climatique renforcée et à une meilleure connaissance des écosystèmes résilients, elle devrait être utilisée pour renforcer ces stratégies de durabilité agricole.

II/ LA BOITE À OUTILS DE L'IA POUR LA PLANÈTE : ATTENTION AUX DÉPENDANCES DE SENTIER

L'utilisation des SIA comme outil au service de l'environnement connaît une forte actualité. De nouvelles publications ont récemment présenté ces promesses au travers d'idées multiples¹⁸⁹. Ces propositions se limitent souvent à des listes de problèmes d'optimisation très précis (ex. optimisation des flux de trafic, optimisation des itinéraires, des réseaux énergétiques intelligents, optimisation de la productivité agricole et de la protection des plantes en agriculture de précision, prévision de la qualité de l'air) et sur des problèmes parfois hérités d'anciens paradigmes organisationnels, urbains, agricoles et sociaux. Bien que cette approche présente un fort potentiel, elle doit être utilisée avec rigueur pour contribuer de manière significative au développement durable. Les publications récentes sur l'IA pour la Terre présentent en effet plusieurs lacunes : l'omission de l'approche du cycle de vie, les risques de « dépendances de sentier » (*path dependancy*), les effets rebonds et l'absence de priorisation en matière d'éco-innovation, ce qui peut entraîner un certain « solutionisme » (résolution locale d'un problème grâce à un outil maîtrisé, mais de manière sous-optimale par manque d'une approche globale et intégrée). Et il n'existe pas de réseau de recherche qui aborde de manière critique les méthodologies de ces interventions.

De façon à utiliser au mieux l'IA pour l'optimisation prédictive de systèmes polluants (transports urbains, énergie de climatisation ou chauffage des bâtiments, agriculture, semences et protection des plantes, gaspillage alimentaire, *smart grids* énergétiques, etc.), huit principes pourraient être adoptés et suivis. Pour illustrer ces principes, prenons le cas d'un projet de SIA pour optimiser les transports urbains, avec un outil de fluidification du trafic automobile :

L'approche de cycle de vie (Iso 14040) pour mesurer les impacts et bénéfices obtenus par ces SIA et anticiper les transferts d'impacts : l'utilisation

massive d'objets connectés et de capteurs en obsolescence programmée pour équiper les voies de circulation entraîne-t-elle de nouveaux impacts sur le cycle de vie (changement climatique, épuisement des ressources, déchets, biodiversité ?)

L'attention aux effets rebonds : si le trafic est plus fluide et qu'il permet un gain de temps en ville, est-ce que certains usagers décideront d'habiter plus loin et donc de polluer plus en participant à l'étalement urbain ?

L'attention aux mécanismes de « dépendance de sentiers » (*path dependency*) : un biais qui conduit à considérer les problèmes toujours de la même manière et à optimiser les infrastructures urbaines existantes avec beaucoup de données disponibles, mais pour de faibles gains environnementaux, tout en retardant la génération d'innovations durables de rupture (ex. un réseau de pistes cyclables et de transport en commun hyper efficace et confortable).

La hiérarchisation de ces SIA selon leur contribution environnementale pour prioriser celles qui apportent des bénéfices environnementaux importants et éviter un « solutionisme » à couleur environnementale : la mise en place de stationnements prédictifs, permettant d'augmenter la probabilité de trouver une place de stationnement dans un quartier à une certaine heure, est-elle une solution prioritaire pour la transition écologique des villes ?

La participation des parties prenantes et des citoyens à la coconstruction des solutions : dans le cas des transports et de la mobilité, les citoyens peuvent aussi contribuer à améliorer des scénarios de mobilité innovants par leurs expériences d'usagers. Une discussion sur la redéfinition des rythmes désirables de la mobilité dans certaines zones pour aborder la coexistence sécuritaire des piétons, vélos, voitures autonomes et véhicules de livraison ne devrait pas se faire que sur la base de données passées, mais aussi sur la possibilité de scénarios prospectifs mis en délibération collective.

¹⁸⁹ Fast. 2017. « 5 Ways Artificial Intelligence Can Help Save The Planet ». En ligne. <https://www.fastcompany.com/40528469/5-ways-artificial-intelligence-can-help-save-the-planet>

World Economic Forum. 2018. « 8 ways AI can help save the planet ». En ligne. <https://www.weforum.org/agenda/2018/01/8-ways-ai-can-help-save-the-planet/>

PwC. 2018. « Fourth Industrial Revolution for the Earth. Harnessing Artificial Intelligence for the Earth ». En ligne. <https://www.pwc.com/gx/en/sustainability/assets/ai-for-the-earth-jan-2018.pdf>

Un répertoire de défis à fort potentiel environnemental pour les SIA, permettant de partager des savoirs et des expériences, devrait être organisé internationalement. Dans notre exemple sur la mobilité, le réseau C40 des villes pionnières sur la lutte contre le changement climatique pourrait organiser une communauté de ce type.

Des politiques de données ouvertes (*open data*) pour les administrations publiques comme pour les entreprises, dès lors que ces données sont d'intérêt général pour la transition énergétique (énergie, déplacements, biodiversité, climat, qualité de l'air, déchets, etc.) permettraient à des acteurs variés de développer des solutions innovantes sur ces défis environnementaux avec un coût limité sur les données.

La littérature numérique sur les données : Iddri et al. (2018 op. cit.) propose aussi de développer une « culture de la donnée » au service de l'écologie par des outils et initiatives pédagogiques pour que tous les acteurs soient capables de lire, créer, exploiter et communiquer des données, notamment les administrations publiques et les collectifs de citoyens.

III/ LA RÉGULATION PRÉDICTIVE DES EFFETS REBONDS : POTENTIEL ET ENJEUX ÉTHIQUES

L'utilisation des SIA dans la régulation algorithmique prédictive des effets rebonds sur les marchés de biens de consommation et d'équipements porte de son côté un fort potentiel pour le développement durable de la société. Ce serait par exemple le cas d'un scénario prospectif où chaque citoyen aurait un crédit de carbone de trois tonnes pour sa consommation annuelle, et serait incité à ne pas dépasser cette limite par des nudges et des recommandations anticipant ses probables effets rebonds (par apprentissage machine supervisé sur des données de comportements de consommation passées).

Mais cette perspective soulève des enjeux éthiques et démocratiques importants : la possible acquisition d'un pouvoir de marché par quelques grandes compagnies ayant la capacité de fournir au système des données environnementales certifiées

à moindre coût par rapport aux PME qui y verraient une barrière à l'entrée ; la non-reconnaissance des initiatives hors marché à fort potentiel pour la transition écologique (ex. comment des initiatives locales d'économie circulaire ou de mobilité durable seront-elle valorisées si elles ne font pas l'objet d'une transaction par le système ?) ; la protection de la vie privée et le pouvoir de normalisation excessif des conduites par les recommandations ; l'absence de processus délibératif sur les recommandations à prioriser. Plusieurs de ces points ont été soulignés lors d'une table de délibération de la coconstruction de la Déclaration de Montréal qui abordait les SIA comme outil de régulation des effets rebonds dans la société.

IV/ L'IA AU SERVICE DE L'INVESTISSEMENT RESPONSABLE

Les SIA sont utilisés en finance de marché pour équiper des dispositifs de « *high-frequency trading* » (HFT) qui sont souvent accusés d'augmenter les risques de krach financier systémique, ou d'accélérer leur propagation, par perte de contrôle des humains.

Les SIA pourraient contribuer à la finance autrement, en renforçant les analyses sur des critères environnementaux et de droits humains pour l'investissement socialement responsable. Un renforcement se faisant par l'apprentissage machine, comme la catégorisation dans les données massives.

Conclusion

Entre le verdissement des SIA et les SIA pour la planète, faut-il choisir ou prioriser l'un par rapport à l'autre pour parvenir à une soutenabilité forte ? Compte tenu de l'urgence de la transition énergétique et écologique, les deux approches devraient être suivies simultanément. La première, parce qu'avec les effets rebonds, il y a des contradictions fortes et non résolues entre les transitions numériques et écologiques. La deuxième, parce qu'elle peut apporter sectoriellement des potentiels d'amélioration significatifs, à condition d'éviter une certaine illusion rhétorique et de suivre les principes que nous avons présentés.

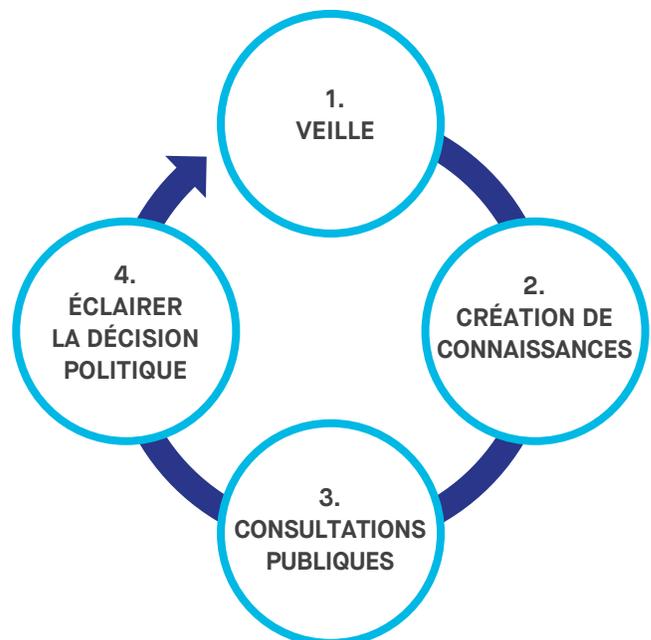
6. LES RECOMMANDATIONS EN VUE DE L'ÉLABORATION DE POLITIQUES PUBLIQUES

Des principes constituant la Déclaration, est élaborée une liste de recommandations dont l'objectif est de proposer des lignes directrices pour réaliser la transition numérique dans le cadre éthique de la Déclaration. Cette liste n'a pas vocation à être exhaustive et ne prétend pas englober tous les secteurs d'application de l'IA ; elle n'inclut pas non plus l'ensemble des recommandations issues de la consultation publique. Il s'agit plutôt de couvrir quelques thèmes intersectoriels clés pour penser la transition vers une société dans laquelle l'IA permet de promouvoir le bien commun : la gouvernance algorithmique, la littératie numérique, l'inclusion numérique de la diversité et la soutenabilité écologique.

Les recommandations qui suivent la Déclaration s'adressent plus spécifiquement aux acteurs du développement de l'IA au Québec et au Canada. Elles constituent des exemples de mesures concrètes élaborées de manière collective à partir des considérations éthiques de la Déclaration. À ce titre, elles peuvent constituer des points de convergence pour les acteurs du développement de l'IA hors du Canada.

RECOMMANDATION 1 : ORGANISME INDÉPENDANT DE VEILLE ET DE CONSULTATION CITOYENNE

Nous recommandons la mise en place d'un organisme de veille et de recherche sur les usages et les impacts sociétaux du numérique et de l'intelligence artificielle. Cet organisme aurait également pour mission de contribuer à l'organisation d'un espace de gouvernance participative en associant les citoyens et les autres parties prenantes, et ainsi d'éclairer les politiques publiques sur la base de la veille, de la production de connaissances et de la participation multipartite.



1.1 Mettre en place un mécanisme de veille continue qui mobiliserait des connaissances sur les aspects techniques, éthiques, juridiques et sociaux du développement des SIA, permettant de surveiller l'émergence de nouveaux enjeux et d'alerter les personnes ressources le cas échéant.

1.1.1 Mobiliser les connaissances interdisciplinaires.

1.1.2 Réaliser une cartographie des meilleures pratiques en matière de gouvernance algorithmique, avec un accent sur les partenariats publics et privés et sur les intérêts en jeu, la pertinence des modèles de *data trust* ou autres mécanismes en lien avec la gestion des communs numériques.

1.1.3 Inclure les associations de citoyens, groupes de réflexion et lanceurs d'alertes susceptibles de mettre en évidence des risques associés au développement des SIA.

1.1.4 Impliquer les différents types de médias sur le numérique et ses impacts, que ce soit pour lancer l'alerte lors de risques identifiés pertinents ou pour le transfert de connaissances auprès du grand public.

1.1.5 Organiser la collecte continue de retours d'expérience sur l'utilisation des SIA dans les organisations publiques et privées, et plus généralement dans la société.

1.2 Favoriser la création de connaissances nouvelles et diversifiées sur les aspects techniques, éthiques, juridiques et sociaux des SIA.

1.2.1 Effectuer de la recherche sur les conditions dans lesquelles les systèmes automatisés publics peuvent contribuer à l'atteinte des objectifs du développement durable.

1.2.2 Créer des appels à projets de recherche innovants favorisant l'interdisciplinarité et la diversité des points de vue (organisme de recherche, organisation de la société civile et parties prenantes).

1.2.3 Produire des rapports bisannuels d'évaluation de la performance des algorithmes publics et de leurs impacts, en accordant une attention particulière aux effets croisés ou cumulés des différents algorithmes sur la situation des individus et des groupes.

1.2.4 Réaliser des projets pilotes à petite échelle, notamment au sein des villes intelligentes, et autres secteurs concernés, afin de déterminer la spécificité des impacts de l'IA dans des contextes donnés.

1.3 Mobiliser les citoyens et les parties prenantes en incluant un volet de consultation proactif qui sera chargé d'évaluer les représentations et les attentes des citoyens à mesure que les SIA se développent, que leurs secteurs d'activité se diversifient et que leur champ d'action s'amplifie.

1.3.1 Sonder les citoyens sur leur perception des enjeux en variant les modalités d'enquête (consultations publiques, groupes de travail, enquête d'opinion en ligne) et en portant une attention particulière à la représentativité sociodémographique des citoyens participants (genre, âge, milieu socioprofessionnel, etc.).

1.3.2 Produire des rapports publics vulgarisés sur les résultats issus des analyses de veille.

1.3.3 Organiser des ateliers de coconstruction qui associent citoyens, organisations de la société civile et parties prenantes, afin d'orienter le développement et le déploiement de l'IA et de proposer des recommandations de politique publique.

1.4 Éclairer la décision publique et amplifier la portée politique des ateliers de coconstruction par un travail d'expertise qui consiste à approfondir les modalités techniques et les recommandations, assurer la cohérence des propositions et produire des brèves et des rapports adressés aux décideurs politiques et aux différents acteurs du développement de l'IA.

RECOMMANDATION 2 : POLITIQUE D'AUDIT ET DE CERTIFICATION DES SIA

Nous recommandons de mettre en place une politique cohérente d'audit et de certification des SIA qui promeut un déploiement responsable (commercialisation, utilisation) des SIA et incite les parties prenantes à adopter les bonnes pratiques pour réduire autant que possible les conséquences néfastes de l'utilisation de SIA et limiter leur utilisation malveillante.

- 2.1 Mettre en place des groupes d'experts multidisciplinaires – soit en utilisant les institutions existantes, soit en les créant ad hoc pour une période déterminée – afin de recenser les ressources institutionnelles et juridiques offrant des capacités de réponses adaptées aux enjeux actuels du déploiement de l'IA et identifier les manques à combler.
- 2.2 Étendre, s'il y a lieu, les compétences des institutions existantes selon leur secteur et leur champ d'action (organisations gouvernementales, organisations d'accréditation, etc.) afin de mettre en œuvre une politique d'audit des algorithmes présentant un risque social élevé, notamment de violation des droits humains, avant leur mise sur le marché et pendant leur exploitation (commerciale ou non).
- 2.3 Étendre les compétences des institutions existantes selon leur secteur et leur champ d'action (organisations gouvernementales, organisations d'accréditations, etc.) afin de délivrer des certifications des SIA qui attestent de la prise en compte des exigences éthiques, sociales et juridiques dans la conception des SIA, et évaluent leurs objectifs de déploiement. La certification devrait être obligatoire pour les SIA utilisés dans les organisations publiques, en particulier gouvernementales (ministères).
- 2.4 Créer une bibliothèque publique, accessible en ligne, des SIA certifiés.
- 2.5 Inciter les entreprises qui développent, commercialisent ou utilisent des SIA à se doter

de comités d'éthique multidisciplinaires et de processus d'audits internes pour identifier les enjeux éthiques, sociaux et juridiques de l'utilisation de SIA dans leurs activités commerciales et leur organisation.

- 2.6 Développer un mécanisme de lanceur d'alerte par la mise en place d'une plateforme en ligne afin de recueillir les informations et les plaintes émises par des individus, des groupes ou des organisations qui suspectent un fonctionnement problématique des SIA.

RECOMMANDATION 3 : ENCAPACITATION ET AUTONOMISATION

Nous recommandons de soutenir l'encapacitation des citoyens face aux technologies du numérique par l'accès à de la formation qui permette la compréhension, la critique, le respect et la responsabilisation afin de participer activement à une société numérique durable.

- 3.1 Promouvoir la littératie numérique par une politique éducative cohérente dans les établissements primaires, secondaires et supérieurs pour développer les compétences de la citoyenneté numérique et former la relève scientifique.
 - 3.1.1 Intégrer l'éducation aux technologies du numérique et de l'intelligence artificielle par l'acquisition de connaissances techniques fondamentales.
 - 3.1.2 Étendre les compétences de la littératie numérique en renforçant l'acquisition des compétences transversales pertinentes pour exercer pleinement la citoyenneté numérique : exploiter des informations et les technologies de l'information, exercer son jugement critique, mettre en œuvre sa pensée créatrice, structurer son identité, etc.
 - 3.1.3 Renforcer l'enseignement éthique relatif aux enjeux du numérique et de l'IA dès les classes de primaire.

3.2 Développer une politique des lieux dédiés à la littératie numérique dans l'espace public afin de faciliter l'accès et l'appropriation de la culture numérique et d'encourager la citoyenneté active, et la diversité d'utilisateurs.

3.2.1 Offrir des espaces de formation, d'expérimentation technologique et d'accueil de la participation citoyenne numérique dans les tiers-lieux comme les bibliothèques publiques, les fab labs, les centres communautaires et culturels.

3.2.2 Prévoir un financement spécifique pour l'acquisition des équipements technologiques nécessaires et la formation du personnel d'encadrement.

3.2.3 Rendre accessibles à tous les formations en faisant un effort particulier pour inclure les groupes isolés ou sous-représentés.

- > Rendre certaines formations mobiles (caravanes des savoirs numériques, boîtes à idées mobiles).
- > Mener une action prioritaire qui cible les groupes sous-représentés (femmes, minorités culturelles, etc.).

3.3 Concevoir une éducation au numérique qui promeut des habitudes de vie favorisant l'autonomie et la santé mentale et physique tout au long de la vie.

3.3.1 Alerter sur les risques de dépendance aux outils numériques, en sensibilisant notamment à l'importance de préserver des moments et espaces de déconnexion.

3.3.2 Entretenir le développement de compétences non numériques comme la capacité à s'orienter sans GPS, savoir écrire à la main, etc.

3.4 Créer une plateforme en accès libre en ligne pour les professionnels de l'éducation, les apprenants, les parents ou tuteurs, et les décideurs publics afin de faciliter la mise

à niveau des connaissances sur les enjeux techniques, éthiques, sociaux et juridiques des technologies du numérique et de l'IA. Cette plateforme servira notamment à :

3.4.1 Répertorier les organisations de l'écosystème de la littératie numérique (établissements d'enseignement, centres de formation, tiers-lieux, entreprises) et coordonner la mobilisation de communautés de pratique dans cet écosystème.

3.4.2 Aider à l'orientation des apprenants, quels que soient leur niveau, leur âge, leurs intérêts.

3.4.3 Constituer une base de connaissances communes sur le numérique et l'IA.

RECOMMANDATION 4 : FORMATIONS EN ÉTHIQUE

Nous recommandons de repenser la formation des parties concernées par la conception, le développement et l'exploitation des SIA en investissant dans la pluridisciplinarité et l'éthique.

4.1 Cibler prioritairement la formation des techniciens de l'IA (les ingénieurs, informaticiens et concepteurs)

4.1.1 Engager, en concertation avec les différentes parties prenantes, une refonte des programmes de formation en génie afin d'intégrer des compétences en éthique, en sciences sociales et en droit afin que les professionnels acquièrent les bons réflexes intellectuels, qu'ils soient sensibilisés aux conséquences potentiellement néfastes des technologies qu'ils développent, et qu'ils élaborent des solutions créatives, éthiquement acceptables et socialement responsables.

4.1.2 Promouvoir une formation éthique et sociale continue afin de faire évoluer les pratiques de conception et développement et d'entretenir la vigilance sur les effets indésirables non prévus des SIA développés.

4.2 Étendre la formation aux utilisateurs de SIA dans le cadre de l'exercice de leur profession et aux gestionnaires qui décident de l'adoption de SIA dans leur organisation.

4.2.1 S'assurer que les professionnels utilisant des SIA comprennent les différents aspects de leur responsabilité, comme le fait de pouvoir justifier une décision prise par le SIA utilisé ou fondée sur une recommandation algorithmique quand cela a un impact personnel ou social important.

4.2.2 Entretenir leur vigilance quant aux éventuelles conséquences éthiques, juridiques et sociales non désirables du SIA utilisé.

4.2.3 Sensibiliser les gestionnaires et les partenaires sociaux aux conséquences de la transition numérique dans leur organisation et les outiller pour procéder à des restructurations socialement responsables.

5.1.1 Tester les SIA sur différentes populations cibles afin d'étudier leurs impacts et déceler les différences de traitement.

5.1.2 Identifier les étiquetages choisis dans les systèmes d'acquisition et d'archivage des données (SAAD), notamment les bases de données qui servent à l'entraînement des SIA, et les paramètres guidant les décisions prises par les SIA publics.

5.1.3 Évaluer la pertinence et l'impact d'un paramètre aléatoire pour les algorithmes de classement (moteurs de recherche et de recommandation), afin de réduire l'importance des bulles de filtre et des biais inéliminables, et d'assurer une diversité de recommandations qui ne reflètent pas les biais de l'algorithme utilisé.

5.1.4 S'assurer que les bases de données d'entraînement utilisées par les SIA publics comprennent un échantillon représentatif des populations concernées.

RECOMMANDATION 5 : FAVORISER UN DÉVELOPPEMENT INCLUSIF DE L'IA

Nous recommandons de mettre en œuvre une stratégie cohérente qui utilise les différentes ressources institutionnelles existantes afin de favoriser un développement inclusif de l'IA et de prévenir les biais et les discriminations potentiels liés au développement et au déploiement des SIA.

5.1 Établir une grille des standards techniques d'inclusion et de non-discrimination dans le fonctionnement des SIA publics et privés. Cette grille doit être unique, évolutive et faire l'objet d'une concertation des différentes organisations habilitées à émettre des réglementations et des normes professionnelles (ministères, ordres professionnels, associations professionnelles). Parmi les dispositions à mettre en œuvre, nous recommandons de :

5.2 Intégrer dans la certification des SIA l'évaluation de leur performance d'inclusivité ou de non-discrimination.

5.3 Investir dans des programmes pour renforcer les compétences en matière d'IA auprès des groupes traditionnellement sous-représentés dans le domaine, notamment auprès des femmes, afin de rendre possible leur inclusion dans toutes les étapes du développement, de la conception à l'application des technologies d'IA.

RECOMMANDATION 6 : PROTÉGER LA DÉMOCRATIE DES MANIPULATIONS POLITIQUES DE L'INFORMATION

Nous recommandons la mise en œuvre d'une stratégie d'endiguement des informations destinées à tromper les citoyens et de la manipulation politique sur les plateformes sociales et les sites internet malveillants, ainsi qu'une stratégie de lutte contre le profilage politique afin de préserver les conditions d'un fonctionnement sain des institutions démocratiques et de l'exercice éclairé de la citoyenneté.

- 6.1 Organiser aux différents niveaux de coordination (provincial, fédéral et international) une conférence des parties prenantes du secteur de l'information et de la communication (sites d'information, réseaux sociaux), des organisations de la société civile du secteur, des décideurs politiques et des citoyens pour mettre en place des standards de certification de l'information et de détection des informations trompeuses.
- 6.2 Inciter les différents sites d'information et les agences de presse dont ils dépendent à créer un organisme commun de vérification des faits (*fact-checking*), à l'échelon provincial, fédéral et international, afin d'améliorer et d'accélérer la vérification des informations, de ne pas entrer dans un marché concurrentiel de la vérification, d'organiser un travail non partisan de vérification et d'augmenter la confiance du public dans l'information.
- 6.3 Favoriser la détection et la signalisation par les internautes des informations trompeuses et des comptes frauduleux en incitant l'organisme commun de vérification, ainsi que les plateformes internet (sites d'information, réseaux sociaux), à proposer à leurs utilisateurs des outils de lancement d'alerte.
- 6.4 Adopter une signalétique commune pour identifier le degré de véracité des informations en ligne, sur la base des standards de certification de l'information.

6.5 Développer des SIA publics de détection des sources d'informations trompeuses sur les plateformes internet et encourager ces plateformes à développer leurs propres outils de détection.

6.6 Adopter une stratégie de découragement des actes malveillants et de ralentissement de la propagation des informations trompeuses, en évitant que les mesures mises en œuvre constituent une censure des opinions politiques qui déplaisent.

6.6.1 Fermer systématiquement des comptes de bots qui propagent des informations trompeuses.

6.6.2 Tarir les sources de revenu publicitaire des sites malveillants et des réseaux sociaux qui ne prennent pas de mesure adéquate pour endiguer la propagation d'informations trompeuses.

RECOMMANDATION 7 : DÉVELOPPEMENT INTERNATIONAL DE L'IA

Nous recommandons l'adoption d'un modèle de développement international non prédateur qui vise l'inclusion des différentes régions du globe sans abuser des pays à faible revenu et à revenu intermédiaire (PFR-PRI). Ce modèle ne doit pas exploiter les retards technologiques ou les failles politiques et juridiques pour capter leurs ressources humaines (les données et les personnes qui ont le potentiel de contribuer au développement local de l'IA).

7.1 Lutter contre l'appropriation des données par des entreprises étrangères et assurer la traçabilité internationale des données.

7.2 S'assurer que les chercheurs, experts et décideurs des PFR-PRI participent activement et équitablement aux discussions internationales sur la régulation de l'IA.

- 7.3 Soutenir les capacités des PFR-PRI à développer leur propre infrastructure numérique et à protéger les données de leur population.
- 7.4 Créer un fonds mondial réservé au renforcement des capacités de « centres d'excellence » de l'IA dans les PFR-PRI, et investir dans des programmes de recherche pour guider la conception, le développement et le déploiement de l'IA.
- 7.5 Soutenir la coopération internationale par des programmes d'échange de chercheurs et d'étudiants entre les pays en pointe dans le développement de l'IA et les pays dont les capacités d'investissement et de développement sont plus réduites.

RECOMMANDATION 8 : EMPREINTE ENVIRONNEMENTALE DIRECTE ET INDUITE DES SIA

Nous recommandons de mettre en œuvre une stratégie publique/privée pour que le développement et le déploiement de SIA et des autres objets numériques soient à la fois compatibles avec une soutenabilité écologique forte et apportent des solutions à la crise environnementale.

- 8.1 Développer une politique d'information et de sensibilisation sur les enjeux de la transition numérique soutenable.
 - 8.1.1 Réaliser et rendre accessibles des bilans environnementaux des SIA pour que leurs impacts sur leur cycle de vie soient connus, compris et pris en compte dans les décisions d'achat et d'investissement.
 - 8.1.2 Diffuser une information pédagogique permettant aux organisations publiques et privées de piloter leur transition numérique de façon soutenable, avec une attention particulière aux effets rebonds et à l'obsolescence programmée des équipements.

8.1.3 Diffuser une information pédagogique permettant aux citoyens d'adopter des styles de vie allant dans le sens d'une vie numérique à très faible impact.

8.1.4 Promouvoir une culture techno-créative et favoriser l'acquisition des compétences permettant de réparer et prolonger la durée de vie des objets et équipements électroniques.

8.2 Élaborer des référentiels d'écoconception des infrastructures et des services des SIA.

8.2.1 Promouvoir dans les entreprises de développement informatique des démarches systématiques d'écoconception des SIA prenant en compte les impacts sur l'ensemble de leur cycle de vie et les risques d'effets rebonds.

8.2.2 Généraliser les démarches d'écoconception des centres de données et des équipements (Internet des objets, capteurs et terminaux utilisateurs de SIA) pour minimiser leur consommation d'énergie et pour prolonger leur durée de vie dans une logique d'économie circulaire.

8.2.3 Développer des SIA et des SAAD (centres de données) favorisant l'utilisation systématique d'une électricité verte (énergies renouvelables, décarbonées) aux différentes étapes de leur cycle de vie, sans détourner cette énergie verte de besoins prioritaires et vitaux pour les populations locales.

8.3 S'engager envers des politiques publiques environnementales ambitieuses pour répondre à l'urgence environnementale.

8.3.1 Définir des politiques publiques pour soutenir la recherche et développement de technologies numériques (Internet des objets, réseaux, centres de données, terminaux) à très faible consommation d'énergie et très faible empreinte environnementale.

8.3.2 Mettre en œuvre un plan d'économie circulaire pour diminuer le besoin d'extraction de ressources naturelles rares utilisées par l'industrie des SIA et mieux gérer les flux de déchets électriques et électroniques.

8.3.3 Alerter les réseaux d'experts internationaux sur l'environnement et le climat pour qu'ils développent spécifiquement des connaissances sur les contradictions les plus urgentes entre la transition écologique et la transition numérique qui est accélérée par l'IA.

8.4 Développer et déployer les SIA comme nouvelle gamme d'outils pour soutenir la transition écologique.

8.4.1 Soutenir l'utilisation des SIA pour accroître la connaissance prédictive des enjeux environnementaux et sociaux, dans une logique de données ouvertes, en donnant la priorité aux enjeux du changement climatique, de la perte de biodiversité, de l'épuisement des ressources, de la qualité de l'air et de l'eau, notamment dans les grandes villes, et des données sur la biomasse et les semences dans des contextes de stress climatique.

8.4.2 Soutenir le développement et le déploiement des SIA pour l'optimisation prédictive de systèmes ayant un impact environnemental (démarches dites d'« IA pour la planète ») comme les enjeux du transport, de la climatisation ou le chauffage des bâtiments, de l'agriculture et de la protection des plantes, de la lutte contre les gaspillages alimentaires, des réseaux énergétiques, avec une attention particulière aux risques de dépendance de sentier et d'effets rebonds.

8.4.3 Expérimenter les SIA comme outil de régulation prédictive des effets rebonds pour établir un système d'incitation à une consommation durable, compatible avec le respect de la vie privée et de la liberté de choix, en portant une attention particulière à la diversité des options documentées dans le dispositif.

8.4.4 Utiliser les SIA au service de l'investissement socialement responsable, quand cela est pertinent, pour calculer l'empreinte carbone, l'empreinte sociale et environnementale des entreprises et des institutions sur leur cycle de vie, et aider à une prise de décision financière orientée vers le développement soutenable.

CRÉDITS DU RAPPORT FINAL

Le rapport de la Déclaration de Montréal IA responsable a été rédigé sous la direction de :

Marc-Antoine Dilhac, instigateur du projet et responsable du Comité d'élaboration de la Déclaration ; codirecteur scientifique de la coconstruction ; professeur au Département de philosophie de l'Université de Montréal ; chaire de recherche du Canada en Éthique publique et théorie politique ; directeur de l'axe Éthique et politique, Centre de recherche en éthique

Christophe Abrassart, codirecteur scientifique de la coconstruction, professeur à l'École de design et codirecteur du Lab Ville Prospective à la Faculté de l'aménagement de l'Université de Montréal, membre du Centre de recherche en éthique

Nathalie Voarino, coordonnatrice scientifique de l'équipe de la Déclaration, candidate au doctorat en bioéthique, Université de Montréal

Coordination

Anne-Marie Savoie, conseillère, vice-rectorat à la recherche, à la découverte, à la création et à l'innovation de l'Université de Montréal

Collaboration aux contenus

Camille Vézy, candidate au doctorat en communication, Université de Montréal

Révision et édition

Chantal Berthiaume, gestionnaire de contenu et rédactrice

Anne-Marie Savoie, conseillère, vice-rectorat à la recherche, à la découverte, à la création et à l'innovation de l'Université de Montréal

Joliane Grandmont-Benoit, coordonnatrice de projets, vice-rectorat aux affaires étudiantes et aux études, Université de Montréal

Traduction

Rachel Anne Normand et François Girard, Services linguistiques Révidaction

Graphisme

Stéphanie Hauschild, directrice artistique

La rédaction de ce rapport n'aurait pu être possible sans les réflexions des citoyens, des professionnels et des experts ayant participé aux ateliers.

NOS PARTENAIRES

Université 
de Montréal



CENTRE DE RECHERCHE EN ETHIQUE



ICRA
Programme
IA et
société



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



