< >

Déclaration de Montréal
IA responsable_

</ >

RAPPORT DE LA DÉCLARATION DE MONTRÉAL POUR UN DÉVELOPPEMENT RESPONSABLE DE L'INTELLIGENCE ARTIFICIELLE

2018

TABLE DES MATIÈRES

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

< >

Déclaration de Montréal

IA responsable_

</ >

LA DÉCLARATION DE MONTRÉAL POUR UN DÉVELOPPEMENT RESPONSABLE DE L'INTELLIGENCE ARTIFICIELLE

2018

TABLE DES MATIÈRES

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

LIRE LA DÉCLARATION DE MONTRÉAL IA RESPONSABLE

UNE DÉCLARATION, POUR QUOI FAIRE?

La Déclaration de Montréal pour un développement responsable de l'intelligence artificielle poursuit trois objectifs :

1. Élaborer un cadre éthique pour le développement et le déploiement de l'IA
2. Orienter la transition numérique afin que tous puissent bénéficier de cette révolution technologique
3. Ouvrir un espace de dialogue national et international pour réussir collectivement un développement inclusif, équitable et écologiquement soutenable de l'IA

UNE DÉCLARATION DE QUOI?

DES PRINCIPES

Le premier objectif de la Déclaration consiste à identifier les principes et les valeurs éthiques qui promeuvent les intérêts fondamentaux des personnes et des groupes. Ces principes appliqués au domaine du numérique et de l'intelligence artificielle restent généraux et abstraits. Pour les lire adéquatement, il convient de garder à l'esprit les points suivants :

> Bien qu'ils soient présentés sous forme de liste, ils ne sont pas hiérarchisés. Le dernier principe n'est pas moins important que le premier. Mais il est possible, selon les circonstances, d'attribuer plus de poids à un principe qu'à un autre, ou de considérer qu'un principe est plus pertinent qu'un autre.

> Bien qu'ils soient divers, ils doivent faire l'objet d'une interprétation cohérente afin d'éviter tout conflit qui empêche leur application. D'une manière générale, les limites de l'application d'un principe sont tracées par le domaine d'application d'un autre principe.

> Bien qu'ils reflètent la culture morale et politique de la société dans laquelle ils ont été élaborés, ils constituent une base pour un dialogue interculturel et international.

> Bien qu'ils puissent être interprétés de diverses manières, ils ne peuvent pas être interprétés de n'importe quelle manière. Il est impératif que l'interprétation soit cohérente.

> Bien que ce soit des principes éthiques, ils peuvent être traduits en langage politique et interprétés de manière juridique.

De ces principes ont été élaborées des recommandations dont l'objectif est de proposer des lignes directrices pour réaliser la transition numérique dans le cadre éthique de la Déclaration. Elles couvrent 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.

UNE DÉCLARATION POUR QUI?

La Déclaration de Montréal est adressée à toute personne, toute organisation de la société civile et toute compagnie désireuses de participer au développement de l'intelligence artificielle de manière responsable, que ce soit pour y contribuer scientifiquement et technologiquement, pour développer des projets sociaux, pour élaborer des règles (règlements, codes) qui s'y appliquent, pour pouvoir en contester les orientations mauvaises ou imprudentes, ou encore pour être en mesure de lancer des alertes à l'opinion publique quand cela est nécessaire.

Elle s'adresse également aux responsables politiques, élus ou nommés, dont les citoyens attendent qu'ils prennent la mesure des changements sociaux en gestation, qu'ils mettent en place rapidement les cadres permettant la transition numérique pour le bien de tous, et qu'ils anticipent les risques sérieux que présente le développement de l'IA.

UNE DÉCLARATION SELON QUELLE MÉTHODE?

La Déclaration est issue d'un processus délibératif inclusif qui met en dialogue citoyens, experts, responsables publics, parties prenantes de l'industrie, des organisations de la société civile et des ordres professionnels. L'intérêt de cette démarche est triple :

1. Arbitrer collectivement les controverses éthiques et sociétales sur l'IA
2. Améliorer la qualité de la réflexion sur l'IA responsable
3. Renforcer la légitimité des propositions pour une IA responsable

L'élaboration de principes et des recommandations est un travail de coconstruction qui a impliqué une diversité de participants dans des lieux publics, dans des salles de réunion d'organisations professionnelles, autour de tables rondes d'experts internationaux, dans des bureaux de chercheurs, dans des salles de cours ou en ligne sur internet, toujours avec la même rigueur.

APRÈS LA DÉCLARATION?

Parce que la Déclaration porte sur une technologie qui n'a cessé de progresser depuis les années 1950 et dont le rythme des innovations majeures s'accélère de manière exponentielle, il est essentiel de concevoir la Déclaration comme un document d'orientation ouvert, révisable et adaptable en fonction de l'évolution des connaissances et des techniques, et des retours d'expériences sur l'utilisation de l'IA dans la société. À la fin du processus d'élaboration de la Déclaration, nous sommes arrivés au point de départ d'une conversation ouverte et inclusive sur l'avenir de l'humanité servie par les technologies de l'intelligence artificielle.

PRÉAMBULE

Pour la première fois dans l'histoire de l'humanité, il est possible de créer des systèmes autonomes capables d'accomplir des tâches complexes que l'on croyait réservées à l'intelligence naturelle : traiter de grandes quantités d'informations, calculer et prédire, apprendre et adapter ses réponses aux situations changeantes, et reconnaître et classer des objets. En raison de la nature immatérielle de ces tâches qu'ils réalisent, et par analogie avec l'intelligence humaine, on désigne ces systèmes très divers par le terme général d'intelligence artificielle. L'intelligence artificielle constitue un progrès scientifique et technologique majeur qui peut engendrer des bénéfices sociaux considérables en améliorant les conditions de vie, la santé et la justice, en créant de la richesse, en renforçant la sécurité publique ou en maîtrisant l'impact des activités humaines sur l'environnement et le climat. Les machines intelligentes ne se contentent pas de mieux calculer que les êtres humains, elles peuvent interagir avec les êtres sensibles, leur tenir compagnie et s'occuper d'eux.

Le développement de l'intelligence artificielle présente cependant des défis éthiques et des risques sociaux majeurs. En effet, les machines intelligentes peuvent contraindre les choix des individus et des groupes, abaisser la qualité de vie, bouleverser l'organisation du travail et le marché de l'emploi, influencer la vie politique, entrer en tension avec les droits fondamentaux, exacerber les inégalités économiques et sociales, et affecter les écosystèmes, l'environnement et le climat. Bien qu'il n'y ait pas de progrès scientifique ni de vie sociale sans risque, il appartient aux citoyens de déterminer les finalités morales et politiques qui donnent un sens aux risques encourus dans un monde incertain.

Les bénéfices de l'intelligence artificielle seront d'autant plus grands que les risques liés à son déploiement seront faibles. Or le premier danger que présente le développement de l'intelligence artificielle consiste à donner l'illusion que l'on maîtrise l'avenir par le calcul. Réduire la société à des nombres et la gouverner par des procédures algorithmiques est un vieux rêve qui nourrit encore les ambitions humaines. Mais dans les affaires humaines, demain ressemble rarement à aujourd'hui, et les nombres ne disent pas ce qui a une valeur morale, ni ce qui est socialement désirable.

Les principes de la présente Déclaration sont les directions d'une boussole éthique qui permet d'orienter le développement de l'intelligence artificielle vers des finalités moralement et socialement désirables. Ils offrent aussi un cadre éthique qui permet de promouvoir les droits humains reconnus internationalement dans les domaines concernés par le déploiement de l'intelligence artificielle. Pris dans leur ensemble, les principes formulés posent enfin les bases de la confiance sociale envers les systèmes artificiellement intelligents.

Les principes de la présente Déclaration reposent sur l'idée commune que les êtres humains cherchent à s'épanouir comme êtres sociaux doués de sensations, d'émotions et de pensées, et qu'ils s'efforcent de réaliser leurs potentialités en exerçant librement leurs capacités affectives, morales et intellectuelles. Il incombe aux différents acteurs et décideurs publics et privés, au niveau local, national et international, de s'assurer que le développement et le déploiement de l'intelligence artificielle soient compatibles avec la protection et l'épanouissement des capacités humaines fondamentales. C'est en fonction de cet objectif que les principes proposés doivent être interprétés de manière cohérente, en tenant compte de la spécificité des contextes sociaux, culturels, politiques et juridiques de leur application.

1

PRINCIPE DE **BIEN-ÊTRE**

Le développement et l'utilisation des systèmes d'intelligence artificielle (SIA) doivent permettre d'accroître le bien-être de tous les êtres sensibles.

1. Les SIA doivent permettre aux individus d'améliorer leurs conditions de vie, leur santé et leurs conditions de travail.

2. Les SIA doivent permettre aux individus de satisfaire leurs préférences, dans les limites de ce qui ne cause pas de tort à un autre être sensible.

3. Les SIA doivent permettre aux individus d'exercer leurs capacités physiques et intellectuelles.

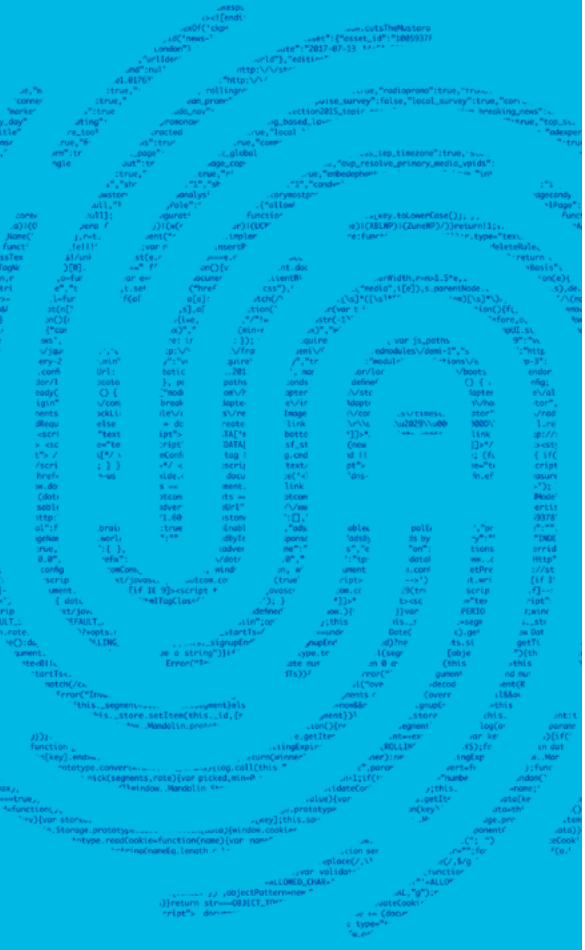
4. Les SIA ne doivent pas constituer une source de mal-être, sauf si ce dernier permet d'engendrer un bien-être supérieur que l'on ne peut atteindre autrement.

5. L'utilisation des SIA ne devrait pas contribuer à augmenter le stress, l'anxiété et le sentiment de harcèlement liés à l'environnement numérique.

2

PRINCIPE DE **RESPECT DE L'AUTONOMIE**

Les SIA doivent être développés et utilisés dans le respect de l'autonomie des personnes et dans le but d'accroître le contrôle des individus sur leur vie et leur environnement.



1. Les SIA doivent permettre aux individus de réaliser leurs propres objectifs moraux et leur conception de la vie digne d'être vécue.

2. Les SIA ne doivent pas être développés ni utilisés pour prescrire aux individus un mode de vie particulier, soit directement, soit indirectement, en mettant en œuvre des mécanismes de surveillance, d'évaluation ou d'incitation contraignants.

3. Les institutions publiques ne doivent pas utiliser les SIA pour promouvoir ni défavoriser une conception de la vie bonne.

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

5. Les SIA ne doivent pas être développés pour propager des informations peu fiables, des mensonges et de la propagande, et devraient être conçus dans le but d'en réduire la propagation.

6. Le développement des SIA doit éviter de créer des dépendances par les techniques de captation de l'attention et par l'imitation de l'apparence humaine qui induit une confusion entre les SIA et les humains.

3

PRINCIPE DE **PROTECTION DE L'INTIMITÉ ET DE LA VIE PRIVÉE**

La vie privée et l'intimité doivent être protégées de l'intrusion de SIA et de systèmes d'acquisition et d'archivage des données personnelles (SAAD).

1. Des espaces d'intimité dans lesquels les personnes ne sont pas soumises à une surveillance, ou à une évaluation numérique, doivent être protégés de l'intrusion de SIA ou de systèmes d'acquisition et d'archivage des données personnelles (SAAD).
2. L'intimité de la pensée et des émotions doit être strictement protégée de l'usage de SIA et de SAAD susceptible de faire du tort, en particulier de l'usage visant à juger moralement des personnes ou de leur choix de vie.
3. Les personnes doivent toujours avoir le choix de la déconnexion numérique dans leur vie privée et les SIA devraient explicitement offrir le choix de la déconnexion à intervalle régulier, sans inciter à rester connecté.
4. Les personnes doivent avoir un contrôle étendu sur les informations relatives à leurs préférences. Les SIA ne doivent pas construire de profils de préférences individuelles pour influencer le comportement des personnes concernées sans leur consentement libre et éclairé.
5. Les SAAD doivent garantir la confidentialité des données et l'anonymisation des profils personnels.
6. Toute personne doit pouvoir garder un contrôle étendu sur ses données personnelles, en particulier par rapport à leur collecte, usage et dissémination. L'utilisation par des particuliers de SIA et de services numériques ne peut être conditionnée à l'abandon de la propriété de ses données personnelles.
7. Toute personne peut faire don de ses données personnelles aux organismes de recherche afin de contribuer au progrès de la connaissance.
8. L'intégrité de l'identité personnelle doit être garantie. Les SIA ne doivent pas être utilisés pour imiter ni modifier l'apparence physique, la voix et d'autres caractéristiques individuelles dans le but de nuire à la réputation d'une personne ou pour manipuler d'autres personnes.

4

PRINCIPE DE **SOLIDARITÉ**

Le développement de SIA doit être compatible avec le maintien de liens de solidarité entre les personnes et les générations.

1. Les SIA ne doivent pas nuire au maintien de relations humaines affectives et morales épanouissantes, et devraient être développés dans le but de favoriser ces relations et de réduire la vulnérabilité et l'isolement des personnes.

2. Les SIA doivent être développés dans le but de collaborer avec les humains sur des tâches complexes et devraient favoriser le travail collaboratif entre les humains.

3. Les SIA ne devraient pas être mis en œuvre pour remplacer des personnes sur des tâches qui requièrent une relation humaine de qualité, mais devraient être développés pour faciliter cette relation.

4. Les systèmes de santé qui recourent aux SIA doivent prendre en considération l'importance pour les patients des relations avec le personnel médical et la famille.

5. Le développement des SIA ne devrait pas stimuler des comportements cruels avec des robots qui prennent l'apparence d'êtres humains ou d'animaux et semblent agir comme eux.

6. Les SIA devraient permettre d'améliorer la gestion des risques et créer les conditions d'une société de mutualisation des risques individuels et collectifs plus efficace.

5

PRINCIPE DE **PARTICIPATION DÉMOCRATIQUE**

Les SIA doivent satisfaire les critères d'intelligibilité, de justifiabilité et d'accessibilité, et doivent pouvoir être soumis à un examen, un débat et un contrôle démocratiques.

1. Le fonctionnement des SIA qui prennent des décisions affectant la vie, la qualité de la vie ou la réputation des personnes doit être intelligible pour leurs concepteurs.
2. Les décisions des SIA affectant la vie, la qualité de la vie ou la réputation des personnes, devraient toujours être justifiables dans un langage compréhensible aux personnes qui les utilisent ou qui subissent les conséquences de leur utilisation. La justification consiste à exposer les facteurs et les paramètres les plus importants de la décision et doit être semblable aux justifications qu'on exigerait d'un être humain prenant le même type de décision.
3. Le code des algorithmes, publics ou privés, doit toujours être accessible aux autorités publiques compétentes et aux parties prenantes concernées à des fins de vérification et de contrôle.
4. La découverte d'erreurs de fonctionnement des SIA, d'effets imprévus ou indésirables, de failles de sécurité et de fuites de données doit être impérativement signalée aux autorités publiques compétentes, aux parties prenantes concernées et aux personnes affectées par la situation.
5. En vertu de l'exigence de transparence des décisions publiques, le code des algorithmes de décision utilisé par les pouvoirs publics doit être accessible à tous, à l'exception des algorithmes présentant, en cas d'usage détourné, un danger sérieux avec une probabilité élevée.
6. Pour les SIA publics ayant un impact important sur la vie des citoyens, ces derniers devraient avoir la possibilité et la compétence de délibérer sur les paramètres sociaux de ces SIA, leurs objectifs et les limites de leur utilisation.
7. On doit pouvoir s'assurer en tout temps que les SIA font ce pour quoi ils ont été programmés et ce pour quoi ils sont utilisés.
8. Tout utilisateur d'un service devrait savoir si une décision le concernant ou l'affectant a été prise par un SIA.
9. Tout utilisateur d'un service qui recourt à des agents conversationnels doit pouvoir identifier facilement s'il interagit avec un SIA ou une personne.
10. La recherche dans le domaine de l'intelligence artificielle devrait rester ouverte et accessible à tous.

6

PRINCIPE D'**ÉQUITÉ**

Le développement et l'utilisation des SIA doivent contribuer à la réalisation d'une société juste et équitable.

1. Les SIA doivent être conçus et entraînés de sorte à ne pas créer, renforcer ou reproduire des discriminations fondées entre autres sur les différences sociales, sexuelles, ethniques, culturelles et religieuses.

2. Le développement des SIA doit contribuer à éliminer les relations de domination entre les personnes et les groupes fondées sur la différence de pouvoir, de richesses ou de connaissance.

3. Le développement des SIA doit bénéficier économiquement et socialement à tous en faisant en sorte qu'il réduise les inégalités et la précarité sociales.

4. Le développement industriel des SIA doit être compatible avec des conditions de travail décentes, et cela, à toutes les étapes de leur cycle de vie, de l'extraction des ressources naturelles jusqu'à leur recyclage, en passant par le traitement des données.

5. L'activité numérique des utilisateurs de SIA et de services numériques devrait être reconnue comme un travail qui contribue au fonctionnement des algorithmes et créé de la valeur.

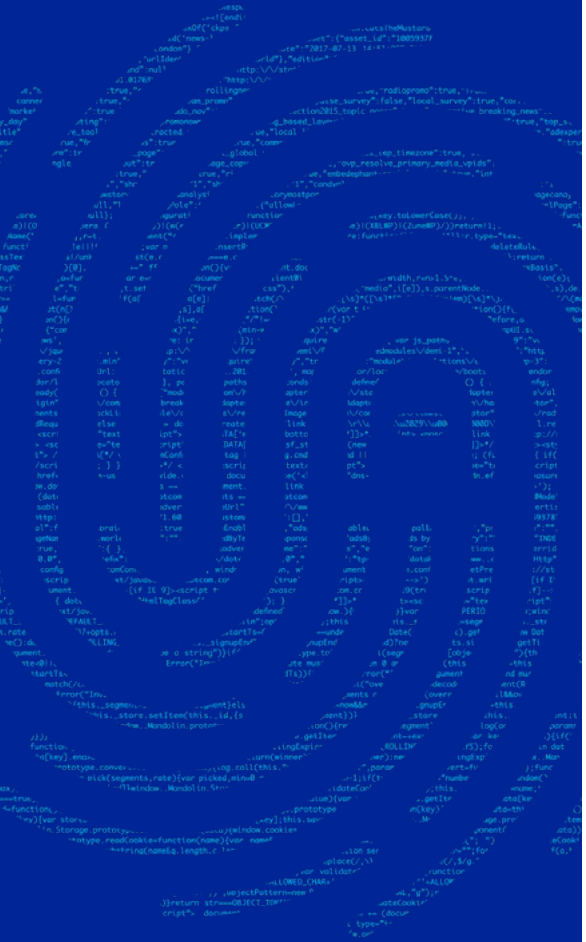
6. L'accès aux ressources, aux savoirs et aux outils numériques fondamentaux doit être garanti pour tous.

7. Le développement de communs algorithmiques et de données ouvertes pour les entraîner et les faire fonctionner est un objectif socialement équitable qui devrait être soutenu.

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.



1. Le développement et l'utilisation de SIA ne devraient pas conduire à une uniformisation de la société par la normalisation des comportements et des opinions.

2. Le développement et le déploiement des SIA doivent prendre en considération les multiples expressions des diversités sociales et culturelles, et cela dès la conception des algorithmes.

3. Les milieux de développement de l'IA, aussi bien dans la recherche que dans l'industrie, doivent être inclusifs et refléter la diversité des individus et des groupes de la société.

4. Les SIA doivent éviter d'enfermer les individus dans un profil d'utilisateur ou une bulle filtrante, de fixer les identités personnelles par le traitement des données de leurs activités passées et de réduire leurs options de développement personnel, en particulier dans les domaines de l'éducation, de la justice et des pratiques commerciales.

5. Les SIA ne doivent pas être utilisés ni développés dans le but de limiter la liberté d'exprimer des idées et de communiquer des opinions, dont la diversité est la condition de la vie démocratique.

6. Pour chaque catégorie de service, l'offre de SIA doit être diversifiée afin que des monopoles de fait ne se constituent pas et ne nuisent aux libertés individuelles.

8

PRINCIPE DE **PRUDENCE**

Toutes les personnes impliquées dans le développement des SIA doivent faire preuve de prudence en anticipant autant que possible les conséquences néfastes de l'utilisation des SIA et en prenant des mesures appropriées pour les éviter.

1. Il est nécessaire de développer des mécanismes qui tiennent compte du potentiel de double-usage (bénéfique et néfaste) de la recherche en IA (qu'elle soit publique ou privée) et du développement des SIA afin d'en limiter les usages néfastes.

2. Lorsque l'utilisation détournée d'un SIA peut représenter un danger sérieux pour la sécurité ou la santé publique, avec une probabilité élevée, il est prudent de restreindre la diffusion publique ou l'accès libre à son algorithme.

3. Avant d'être mis sur le marché, qu'ils soient payants ou gratuits, les SIA doivent satisfaire des critères rigoureux de fiabilité, de sécurité et d'intégrité, et faire l'objet de tests qui ne mettent pas en danger la vie des personnes, ne nuisent pas à leur qualité de vie ni ne portent atteinte à leur réputation ou leur intégrité psychologique. Ces tests doivent être ouverts aux autorités publiques compétentes et aux parties prenantes concernées.

4. Le développement des SIA doit prévenir les risques d'une utilisation néfaste des données d'utilisateurs et protéger l'intégrité et la confidentialité des données personnelles.

5. Les erreurs et les failles découvertes dans les SIA et SAAD devraient être partagées publiquement par les institutions publiques et les entreprises dans les secteurs qui présentent un danger important pour l'intégrité personnelle et l'organisation sociale, et ce, à l'échelle mondiale.

9

PRINCIPE DE **RESPONSABILITÉ**

Le développement et l'utilisation des SIA ne doivent pas contribuer à une déresponsabilisation des êtres humains quand une décision doit être prise.

1. Seuls des êtres humains peuvent être tenus responsables de décisions issues de recommandations faites par des SIA et des actions qui en découlent.

2. Dans tous les domaines où une décision qui affecte la vie, la qualité de la vie ou la réputation d'une personne doit être prise, la décision finale devrait revenir à un être humain et cette décision devrait être libre et éclairée.

3. La décision de tuer doit toujours être prise par des êtres humains et la responsabilité de cette décision ne peut être transférée à un SIA.

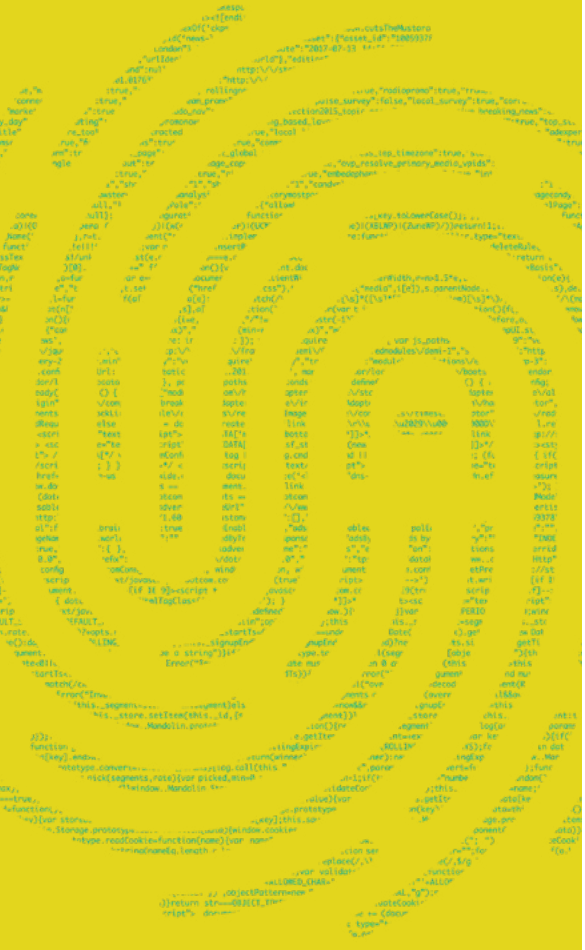
4. Les personnes qui autorisent des SIA à commettre un crime ou un délit, ou qui font preuve de négligence en les laissant en commettre, sont responsables de ce crime ou de ce délit.

5. Dans le cas où un tort a été infligé par un SIA, et que le SIA s'avère fiable et a fait l'objet d'un usage normal, il n'est pas raisonnable d'en imputer la faute aux personnes impliquées dans son développement ou son utilisation.

10

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.



1. Les équipements de SIA, leurs infrastructures numériques et les objets connectés sur lesquels ils s'appuient comme les centres de données, doivent viser la plus grande efficacité énergétique et minimiser les émissions de gaz à effet de serre (GES) sur l'ensemble de leur cycle de vie.

2. Les équipements de SIA, leurs infrastructures numériques et les objets connectés sur lesquels ils s'appuient, doivent viser à générer un minimum de déchets électriques et électroniques et prévoir des filières de maintenance, de réparation et de recyclage dans une logique d'économie circulaire.

3. Les équipements de SIA, leurs infrastructures numériques et les objets connectés sur lesquels ils s'appuient, doivent minimiser les impacts sur les écosystèmes et la biodiversité à toutes les étapes de leur cycle de vie, notamment lors de l'extraction des ressources naturelles et des étapes de fin de vie.

4. Les acteurs publics et privés doivent soutenir le développement de SIA écologiquement responsables afin de lutter contre le gaspillage des ressources naturelles et des biens produits, de mettre en place des chaînes d'approvisionnement et des échanges commerciaux soutenables, et de réduire la pollution à l'échelle planétaire.

LEXIQUE

Activité numérique

On entend par activité numérique l'ensemble des actions posées par un individu dans un environnement numérique, que ce soit sur un ordinateur, un téléphone ou tout autre objet connecté.

Agent conversationnel (*chatbot*)

Un agent conversationnel est un système d'IA qui peut dialoguer avec son utilisateur en langage naturel.

Algorithme

Un algorithme est une méthode de résolution de problèmes par une suite finie et non ambigüe d'opérations. Plus précisément dans le domaine de l'intelligence artificielle, il s'agit de la suite d'opérations appliquées aux données d'entrées pour arriver au résultat désiré.

Apprentissage machine (*machine learning*)

L'apprentissage machine est la branche de l'intelligence artificielle qui consiste à programmer un algorithme à apprendre par lui-même. Parmi la multitude de techniques, on distingue trois types majeurs d'apprentissage machine :

> En apprentissage supervisé, le système d'intelligence artificielle (SIA) apprend à prédire une valeur à partir d'une donnée entrée. Cela nécessite d'avoir des couples entrée-valeur annotés lors de l'entrainement. Par exemple, un système peut apprendre à reconnaître l'objet présent sur une photo.

> En apprentissage non-supervisé, le SIA apprend à trouver des similitudes entre des données qui n'ont pas été annotées, par exemple afin de les diviser en différentes partitions homogènes. Ainsi, un système peut reconnaître des communautés d'utilisateurs de réseaux sociaux.

> En apprentissage par renforcement, le SIA apprend à agir sur son environnement de façon à maximiser une récompense qui lui est donnée lors de l'entrainement. C'est la technique avec laquelle des SIA ont pu battre des humains au jeu de Go ou au jeu vidéo Dota2.

Apprentissage profond (*deep learning*)

L'apprentissage profond est la branche de l'apprentissage machine qui utilise des réseaux de neurones artificiels à plusieurs niveaux. C'est la technologie qui est derrière les plus récentes avancées en IA.

Biens communs numériques (*digital commons*)

Les biens communs numériques sont les applications ou les données produites par une communauté. Contrairement aux biens matériels, ils sont facilement partageables et ne se détériorent pas lorsqu'ils sont utilisés. Ainsi, par opposition aux logiciels propriétaires, les logiciels open source – qui résultent souvent d'une collaboration entre programmeurs – constituent des biens communs numériques puisque leur code source est ouvert, c'est-à-dire accessible à tous.

Bulle de filtre (*filter bubble*)

L'expression bulle de filtre (ou bulle filtrante) désigne l'information « filtrée » qui parvient à un individu lorsqu'il est sur internet. En effet, divers services comme les réseaux sociaux ou les moteurs de recherche offrent des résultats personnalisés à leurs utilisateurs. Ceci peut avoir pour effet d'isoler les individus (dans des « bulles ») puisqu'ils n'accèdent plus à une information commune.

Déconnexion numérique
On entend par déconnexion numérique l'arrêt temporaire ou permanent par un individu de son activité numérique.

Dépendance de sentier
Mécanisme social par lequel des décisions technologiques, organisationnelles ou institutionnelles, jugées rationnelles à une époque mais devenues sous-optimales aujourd'hui, continuent malgré tout d'influencer la prise de décision. Un mécanisme maintenu à cause d'un biais cognitif ou parce que son changement conduirait à un coût ou un effort trop élevé. C'est par exemple le cas des infrastructures routières urbaines lorsqu'elles conduisent à des programmes d'optimisation de la circulation, au lieu d'envisager un changement pour organiser une mobilité à très faibles émissions de carbone. Ce mécanisme doit être connu lors de l'utilisation de l'IA pour des projets sociaux, car les données d'entrainement en apprentissage supervisé peuvent parfois renforcer d'anciens paradigmes organisationnels dont la pertinence fait débat aujourd'hui.

Développement soutenable (*sustainable*)
Le développement soutenable (ou durable) désigne un développement des sociétés humaines qui est compatible avec la capacité des systèmes naturels à offrir les ressources et les services nécessaires à ces sociétés. Il s'agit d'un développement économique et social qui répond aux besoins des personnes actuelles sans compromettre l'existence des générations futures.

Données ouvertes (*open data*)
Les données ouvertes désignent les données numériques auxquelles les usagers peuvent accéder librement. C'est par exemple le cas pour la plupart des résultats de recherche publiés en IA.

Données personnelles
Les données personnelles sont celles qui permettent d'identifier directement ou indirectement un individu.

Effet rebond
L'effet rebond est le mécanisme par lequel une plus grande efficacité énergétique ou une meilleure performance environnementale des biens, équipements et services, conduit à une augmentation plus que proportionnelle de leur consommation. Par exemple, la taille des écrans augmente, la quantité des appareils électroniques augmente dans les ménages, et on parcourt de plus grandes distances en voiture ou en avion. Il en résulte globalement une plus grande pression sur les ressources et l'environnement.

Entrainement
L'entrainement est le processus de l'apprentissage machine pendant lequel le SIA construit un modèle à partir de données. Les performances du SIA dépendront de la qualité du modèle, lui-même dépendant de la quantité et de la qualité des données utilisées durant l'entrainement.

Fiabilité
Un système d'IA est fiable lorsqu'il effectue la tâche pour laquelle il a été conçu de manière attendue. La fiabilité est la probabilité de succès qui varie entre 51% et 100%, c'est-à-dire qui est strictement supérieur au hasard. Plus un système est fiable, plus son comportement est prévisible.

GAN
Acronyme de Generative Adversarial Network, en français Réseaux Antagonistes Génératifs. Dans un GAN, deux réseaux antagonistes sont en compétition pour générer une image. Ils peuvent être par exemple utilisés pour créer une image, un enregistrement ou une vidéo paraissant quasi-réels pour un humain.

Intelligence artificielle (IA)
L'intelligence artificielle (IA) désigne l'ensemble des techniques qui permettent à une machine de simuler l'intelligence humaine, notamment pour apprendre, prédire, prendre des décisions et percevoir le monde environnant. Dans le cas d'un système informatique, l'intelligence artificielle est appliquée à des données numériques.

Intelligibilité

Un système d'IA est intelligible lorsqu'un être humain doté des connaissances nécessaires peut comprendre son fonctionnement, c'est-à-dire son modèle mathématique et les processus qui le déterminent.

Justifiabilité d'une décision

La décision d'un système d'IA est justifiée lorsqu'il existe des raisons non triviales qui motivent cette décision et que ces raisons sont communicables en langage naturel.

Littératie numérique

La littératie numérique d'un individu désigne son habilité à accéder, gérer, comprendre, intégrer, communiquer, évaluer et créer de l'information de façon sécuritaire et appropriée au moyen des outils numériques et des technologies en réseaux pour participer à la vie économique et sociale.

Soutenabilité écologique forte

La notion de soutenabilité (ou durabilité) écologique forte renvoie à l'idée que pour être soutenable, le rythme de consommation des ressources naturelles et d'émissions de polluants doit être compatible avec les limites environnementales planétaires, le rythme de renouvellement des ressources et des écosystèmes, ainsi que la stabilité du climat. Contrairement à la soutenabilité faible, moins exigeante, la soutenabilité forte n'admet pas qu'on substitue des pertes de ressources naturelles par du capital artificiel.

Système d'acquisition et d'archivage des données (SAAD)

Un SAAD désigne tout système informatique pouvant collecter et enregistrer des données. Celles-ci seront éventuellement utilisées pour l'entrainement d'un système d'IA ou comme paramètres pour une prise de décision.

Système d'intelligence artificielle (SIA)

Un système d'IA désigne tout système informatique utilisant des algorithmes d'intelligence artificielle, que ce soit un logiciel, un objet connecté ou un robot.

CRÉDITS

La rédaction de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle est le fruit du travail d'une équipe scientifique pluridisciplinaire et interuniversitaire qui s'appuie sur un processus de consultation citoyenne et sur la concertation avec des experts et des parties prenantes du développement de l'IA.

Christophe Abrassart, professeur agrégé 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 (CRÉ)

Yoshua Bengio, professeur titulaire au Département d'informatique et recherche opérationnelle (DIRO) de l'Université de Montréal, directeur scientifique du Mila et de l'IVADO

Guillaume Chicoisne, directeur des programmes scientifiques, IVADO

Nathalie de Marcellis-Warin, présidente directrice générale du Centre interuniversitaire de recherche en analyse des organisations (CIRANO), professeur titulaire à Polytechnique Montréal

Marc-Antoine Dilhac, professeur agrégé au Département de philosophie de l'Université de Montréal; directeur de l'axe Éthique et politique, Centre de recherche en éthique; directeur de l'Institut Philosophie Citoyenneté Jeunesse; chaire de recherche du Canada en Éthique publique et théorie politique

Sébastien Gambs, professeur d'informatique à l'UQAM, Chaire de recherche du Canada en analyse respectueuse de la vie privée et éthique des données massives

Vincent Gautrais, professeur titulaire à la Faculté de droit de l'Université de Montréal; directeur du Centre de recherche en droit public (CRDP)

Martin Gibert, conseiller en éthique pour IVADO et chercheur au Centre de recherche en éthique

Lyse Langlois, professeure titulaire et vice-doyenne à la recherche, Faculté des sciences sociales, Département des relations industrielles, Université Laval; directrice de l'Institut d'éthique appliquée (IDÉA), chercheure au Centre interuniversitaire sur la mondialisation et le travail (CRIMT)

François Laviolette, professeur titulaire au Département d'informatique et de génie logiciel de l'Université Laval, directeur du Centre de recherche en données massives (CRDM)

Pascale Lehoux, professeur titulaire à l'École de santé publique de l'Université de Montréal (ESPUM); Chaire de l'Université de Montréal sur l'innovation responsable en santé

Jocelyn Maclure, professeur titulaire à la Faculté de philosophie à l'Université Laval; président de la Commission de l'éthique en science et technologie (CEST)

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

Joëlle Pineau, professeure agrégée à la School of Computer Science de l'Université McGill, directrice du Laboratoire IA de Facebook, codirectrice du Laboratoire Reasoning and Learning

Peter Railton, Gregory S. Kavka Distinguished University Professor; John Stephenson Perrin Professor; Arthur F. Thurnau Professor, au département de philosophie de l'Université du Michigan et membre de l'Académie américaine des arts et des sciences

Catherine Régis, professeure agrégée à la Faculté de droit de l'Université de Montréal; Chaire de recherche du Canada sur la culture collaborative en droit et politiques de la santé; chercheuse régulière, Centre de recherche en droit public (CRDP)

Christine Tappolet, professeure titulaire au Département de philosophie de l'Université de Montréal, directrice du Centre de recherche en éthique (CRÉ), responsable du Groupe interuniversitaire sur la normalité (GRIN)

Nathalie Voarino, coordonnatrice scientifique de la Déclaration, candidate au doctorat en Sciences biomédicales, option Bioéthique, Université de Montréal

< >

Déclaration de Montréal
IA responsable_

</ >

PARTIE 1

DÉMARCHE ET MÉTHODOLOGIE

TABLE DES MATIÈRES

RÉDACTION

MARC-ANTOINE DILHAC, codirecteur scientifique de la coconstruction, professeur au Département de philosophie de l'Université de Montréal ; directeur de l'axe Éthique et politique, Centre de recherche en éthique ; chaire de recherche du Canada en Éthique publique et théorie politique

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 (CRÉ)

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

TABLE DES MATIÈRES

TABLE DES TABLEAUX ET DES FIGURES

SOMMAIRE

Le 3 novembre 2017, l'Université de Montréal lançait les travaux de coconstruction de la *Déclaration de Montréal* pour un développement responsable de l'intelligence artificielle (*Déclaration de Montréal*). Un an plus tard, nous présentons les résultats du processus de délibération citoyenne. Des dizaines d'événements ont été organisés pour engager la discussion autour des enjeux sociétaux de l'intelligence artificielle (IA), et une quinzaine d'ateliers de délibération ont été tenus en trois mois, faisant participer plus de 500 citoyens, experts et parties prenantes de tous les horizons.

La *Déclaration de Montréal* est une œuvre collective qui a pour objectif de mettre le développement de l'IA au service du bien-être de tout un chacun, et d'orienter le changement social en élaborant des recommandations ayant une forte légitimité démocratique.

La méthode retenue de la coconstruction citoyenne s'appuie sur une déclaration préliminaire de principes éthiques généraux qui s'articulent autour de 7 valeurs fondamentales : l'autonomie, la justice, le bien-être, la vie privée, la démocratie, la connaissance et la responsabilité. À la fin du processus, la Déclaration a été enrichie et présente désormais 10 principes autour des valeurs suivantes : le bien-être, l'autonomie, l'intimité et la vie privée, la solidarité, la démocratie, l'équité, l'inclusion, la prudence, la responsabilité et la soutenabilité environnementale.

Si l'un des objectifs du processus de coconstruction est d'affiner les principes éthiques proposés dans la version préliminaire de la *Déclaration de Montréal*, un autre objectif tout aussi important consiste à élaborer des recommandations pour encadrer la recherche en IA et son développement technologique et industriel.

D'abord, qu'est-ce que l'IA ?

Très brièvement, l'IA consiste à simuler certains processus d'apprentissage de l'intelligence humaine, à s'en inspirer et à les reproduire. Par exemple, découvrir des motifs complexes parmi une grande quantité de données, ou encore raisonner de manière probabiliste, afin de classer en fonction de catégories des informations, de prédire une donnée quantitative ou de regrouper des données ensemble. Ces compétences cognitives sont à la base d'autres compétences comme celles de décider entre plusieurs actions possibles pour atteindre un objectif, d'interpréter une image ou un son, de prédire un comportement, d'anticiper un événement, de diagnostiquer une pathologie, etc. Ces réalisations de l'IA reposent sur deux éléments : des données et des algorithmes, c'est-à-dire des suites d'instructions permettant d'accomplir une action complexe.

Pour discuter concrètement des enjeux éthiques de l'IA, la **méthode des ateliers de coconstruction** s'appuie sur la version préliminaire de la *Déclaration de Montréal*. Schématiquement, après avoir statué sur le « quoi ? » (quels principes éthiques souhaitables devraient être rassemblés dans une déclaration sur l'éthique de l'IA ?), il s'agit d'anticiper par la prospective, avec les participants, comment des enjeux éthiques pourraient surgir dans les prochaines années à propos de l'IA, dans les secteurs de la santé, de la justice, de la ville intelligente, de l'éducation et de la culture, du monde du travail et des services publics. Ensuite, on imagine comment on pourrait répondre à ces enjeux. Par exemple, par un dispositif comme une certification sectorielle, un nouvel acteur-médiateur, un formulaire ou une norme, par une politique publique ou un programme de recherche.

Les citoyens et parties prenantes ont donc participé à des cafés citoyens ou des journées complètes de coconstruction, où ils ont pu débattre autour de scénarios prospectifs.

D'autres citoyens ont choisi de contribuer à la réflexion en répondant à un questionnaire en ligne, ou en déposant un mémoire. Les résultats de ces démarches spécifiques seront rapportés dans le rapport global des activités liées à la *Déclaration de Montréal* à paraître à l'automne 2018.

Les résultats des ateliers de coconstruction – Les grandes orientations

De manière générale, les participants ont reconnu que l'avènement de l'IA s'accompagne d'importants bénéfices potentiels. Notamment, dans le secteur du travail, les participants ont reconnu le gain de temps que pourrait offrir le recours à des dispositifs d'IA. Cependant, il a aussi été soutenu que le développement de l'IA devait se faire avec prudence afin de prévenir les dérives et les mauvais usages.

Les citoyens ont fait ressortir la nécessité de mettre en place différents mécanismes pour assurer la qualité, l'intelligibilité, la transparence et la pertinence des informations transmises. Ils ont également souligné la difficulté à garantir un véritable consentement éclairé.

La grande majorité des participants a reconnu la nécessité d'aligner les intérêts privés avec les intérêts publics et d'empêcher l'apparition de monopoles, voire de limiter l'influence de grandes entreprises.

Les participants recommandent également de mettre en place des mécanismes qui émaneraient et impliqueraient des personnes indépendantes et formées aux enjeux technologiques et éthiques du numérique et de l'IA, afin de favoriser la diversité et l'intégration des plus vulnérables, et de protéger la pluralité des modes de vie.

Quelles que soient les applications, la majorité des participants souligne le fait que l'IA doit rester un outil et que la décision finale doit rester celle d'un humain quand des intérêts fondamentaux sont en jeu.

Les enjeux prioritaires, en fonction des principes de la *Déclaration de Montréal*

Le principe de responsabilité a été jugé l'enjeu prioritaire, suivi de ceux de respect de l'autonomie, de protection de la vie privée. Viennent ensuite ceux de promotion du bien-être, de connaissance et de justice. Il faut cependant noter qu'ils sont tous étroitement liés.

Pour ce qui est du principe d'autonomie, qu'une majorité de participants considèrent comme prioritaire, il a trait au respect et à la promotion de l'autonomie individuelle face à des risques de contrôle par les technologies et de dépendance aux outils. Il soulève également l'enjeu d'une double liberté de choix : pouvoir suivre son propre choix face à une décision orientée par l'IA, mais également pouvoir choisir de ne pas utiliser ces outils sans pour autant risquer une exclusion sociale.

Le principe de bien-être occupe également une place importante pour les participants. Il est présent en filigrane à toutes les tables, manifestant un souhait collectif d'avancer vers une société juste, équitable et favorisant le développement de tous. De façon générale, le principe de bien-être a également pris la forme d'un appel au maintien d'une relation humaine et émotionnelle de qualité entre experts et usagers dans tous les secteurs.

Des enjeux ayant conduit à la création de nouveaux principes, ou de nouveaux thèmes à explorer et à délibérer

Le principe de justice a été abordé selon deux types d'enjeux, ce qui pourrait donner lieu à **2 nouveaux principes** :

1. un **principe de diversité** visant à éviter les discriminations en trouvant des mécanismes dépourvus de biais
2. un **principe d'équité** ou de justice sociale, impliquant que les bénéfices de l'IA soient accessibles à tous, et que le développement de l'IA ne contribue pas à l'accroissement des inégalités économiques et sociales, mais qu'il les réduise

Un **principe de prudence**. Les enjeux relatifs à la confiance envers le développement des technologies de l'IA ont régulièrement été soulevés. Cet enjeu de confiance entretient ainsi une relation étroite avec la question de la fiabilité des systèmes de l'IA.

Un **principe d'explicabilité ou de justifiabilité**. Ce principe implique de pouvoir comprendre une décision algorithmique et agir face à elle. Pour cela, les citoyens accordent de l'importance à l'explicabilité des procédures algorithmiques afin de pouvoir comprendre et vérifier quels critères ont été pris en compte dans la décision.

Un **principe de soutenabilité environnementale**. L'impact du développement et de l'utilisation de l'IA sur l'**environnement** soulèvent des enjeux particuliers, notamment la façon de garantir l'utilisation responsable et équitable des ressources matérielles et naturelles.

Les mécanismes pour la transition numérique

Toutes les tables de coconstruction se sont entendues sur **3 mécanismes** prioritaires pour garantir un développement socialement responsable de l'IA et ce, quel que soit le secteur :

1. Des dispositions légales
2. La mise en place de formations pour tous
3. L'identification d'acteurs clés et indépendants pour la gestion de l'IA

Poursuivre la délibération

Les travaux de la *Déclaration de Montréal* se sont concentrés dans cette première année de consultation sur plusieurs secteurs clés : éducation, santé, travail, ville intelligente et police prédictive, environnement, démocratie et propagande des médias. Il est évident qu'une année de coconstruction ne permet pas de couvrir tous les enjeux éthiques et sociétaux de l'IA. La *Déclaration de Montréal* n'est pas seulement le fruit d'un processus de réflexion collective, elle est ce processus lui-même : au-delà de la Déclaration Montréal An 1, le processus de consultation et de réflexion collective se poursuit parce que l'évolution technologique n'attend pas.

C'est autour de chantiers prioritaires que nous présentons les recommandations de politique publique. À ce jour, nous pouvons affirmer que 4 chantiers se sont imposés : gouvernance algorithmique, littératie numérique, diversité et inclusion, et transition écologique.

1. INTRODUCTION

Le 3 novembre 2017, l'Université de Montréal, en collaboration avec les Fonds de recherche du Québec, lançait les travaux de coconstruction de la **Déclaration de Montréal pour un développement responsable de l'intelligence artificielle** (*Déclaration de Montréal*). Nous n'anticipions pas l'intérêt qu'allait susciter cette démarche, ni l'ampleur de la tâche qui nous attendait. Un an plus tard, nous présentons les résultats du processus de délibération qui a impliqué divers groupes de la société civile, citoyens, experts, ordres professionnels, parties prenantes de l'industrie, décideurs publics. Ce bilan est très fructueux : des dizaines d'événements ont été organisés pour engager la discussion autour des enjeux sociétaux de l'IA, et une quinzaine d'ateliers de délibération ont été tenus sur une période de février à octobre, faisant participer plus de 500 citoyens, experts et parties prenantes de tous les horizons professionnels.

Le rapport que nous présentons doit se lire comme le bilan d'un exercice de délibération démocratique pour éclairer les choix de politique publique en matière d'intelligence artificielle, exercice qui peut servir de référence pour d'autres exercices délibératifs à venir. Les travaux de ce que l'on appelle la *Déclaration de Montréal* ont été menés par une communauté pluridisciplinaire et interuniversitaire de chercheurs, principalement au Québec mais également dans le reste du monde. La prise de conscience face aux enjeux sociétaux de l'intelligence artificielle est partagée par cette communauté de recherche, mais elle l'est aussi dans l'ensemble de la société. Nous avons proposé une démarche de coconstruction citoyenne parce que nous avons la conviction que tout le monde a son mot à dire sur l'organisation de notre société. Cette démarche est innovante dans son contenu et dans sa conduite : tout d'abord, elle met en œuvre une conception prospective de l'éthique appliquée, qui consiste à anticiper les controverses éthiques sur des technologies d'intelligence artificielle en devenir ou sur des situations sociales où ces technologies sont utilisées de manière inédite. Ensuite, nous avons conduit cette démarche de consultation avec une ampleur, elle aussi inédite. Les chiffres cités plus haut l'indiquent clairement. Cette démarche devrait se poursuivre au-delà de la présentation publique de la *Déclaration de Montréal* dans la mesure où elle devra rester un objet ouvert à la révision.

Si nous avons interpelé le public autour de la rédaction de la Déclaration, nous avons en retour été interpelés par le public et aussi par les parties prenantes : que peut changer la Déclaration ? Qui la rédige ? N'est-ce pas un exercice d'universitaires un peu vain ? N'y-t-il pas déjà trop de manifestes, de professions de foi sur les valeurs éthiques de l'intelligence artificielle ? Vouloir encadrer avec des principes éthiques et des recommandations le développement de l'intelligence artificielle n'est-ce pas finalement l'endosser ? Cela ne revient-il pas à approuver une vision techniciste de la société ? Pourquoi ne pas plutôt consacrer notre

énergie à critiquer ce développement ? Aucune de ces interpellations n'est mauvaise, et parce que nous nous sommes engagés à promouvoir une plus grande transparence de l'intelligence artificielle, nous nous sommes aussi engagés à plus de transparence dans le processus que nous avons mis en place. Ce rapport, nous l'espérons, apportera quelques réponses.

D'éthique de l'intelligence artificielle, il en est beaucoup question depuis deux ans dans différents pays. Tous les acteurs de son développement, chercheurs, entreprises, citoyens, représentants politiques, reconnaissent l'urgence d'établir un cadre éthique, politique et légal pour orienter la recherche et les applications de l'intelligence artificielle. Car il ne fait pas de doute que nous sommes à l'orée d'une nouvelle révolution industrielle avec l'essor des technologies de l'intelligence artificielle. Les impacts de cette révolution sur la production des biens, la prestation de services, l'organisation du travail et du marché de l'emploi, ou encore sur les relations personnelles et familiales sont encore mal connus mais seront très importants, peut-être déstabilisants dans certains secteurs. En effet, les changements sociétaux induits par l'intelligence surprennent par leur soudaineté et suscitent des réactions variées, de l'enthousiasme à la réprobation en passant par le scepticisme. Nous pourrions les ignorer et nous lancer dans des spéculations sur l'existence ou non de ce que l'on appelle l'intelligence artificielle mais nous reporterions alors simplement le problème à un temps où il ne sera plus possible d'agir pour orienter son développement.

De nombreuses objections et craintes ont été exprimées lors de ce premier parcours de coconstruction. Plusieurs participants aux ateliers et plusieurs observateurs des travaux de la Déclaration ont mis en cause l'idéologie techniciste qui voit dans la technologie le moyen d'organiser rationnellement toute la société et qui réduit les enjeux sociaux à des problèmes techniques. D'autres questionnent la capacité et la volonté des institutions publiques à réguler des technologies lucratives. Ces objections ne doivent pas être balayées d'un revers de main parce qu'elles sont fondées sur des précédents historiques qui ont ébranlé la confiance dans les innovations technologiques et plus encore dans leurs promoteurs. Mais il est important que ceux qui formulent ces objections fassent aussi en sorte de ne pas miner tout effort pour orienter positivement l'avenir de notre société et soutiennent, en y participant, la délibération démocratique qui nous permet de garder le contrôle sur le développement des technologies du numérique et de l'intelligence artificielle. On peut déplorer les effets de ces nouvelles technologies sur le lien social, on peut critiquer la réduction de la vie sociale à un ensemble de modes de vie, cela n'arrêtera pas l'innovation technologique, ni ne la fera dévier. Or, c'est tout l'enjeu de la *Déclaration de Montréal* : orienter le développement de l'intelligence artificielle afin de promouvoir des intérêts éthiques et sociétaux fondamentaux et offrir des repères pour protéger les droits humains.

Pour finir, nous ne présentons pas dans ce rapport une théorie de l'intelligence artificielle et nous ne défendons pas non plus d'arguments sophistiqués pour trancher la question lancinante concernant l'usage du terme « intelligence artificielle » : est-ce un terme approprié pour désigner les algorithmes de traitement d'informations, de reconnaissance et de décision ? Certains contestent l'usage de ce terme en opposant le fait que l'intelligence artificielle renvoie à des processus de connaissance très limités en comparaison avec l'intelligence humaine, voire à l'intelligence comportementale des pigeons. C'est indéniable. Mais dans ce cas, il faut aussi reconnaître que la paramécie offre une complexité qui surpasse celle de n'importe quel algorithme, fût-il apprenant. En poursuivant ce chemin, on tombe sur l'impasse de la compréhension de l'intelligence tout court. Qu'est-ce que l'intelligence humaine ? Y en a-t-il une ou plusieurs formes ? Doit-on introduire et spécifier une forme « émotionnelle » de l'intelligence ? Et pourquoi refuser dans ce cas l'introduction d'une forme « artificielle » de l'intelligence ? Les centaines de milliers de pages qui ont été produites pour répondre à ce genre de question n'y suffisent pas.

Cependant, quelques remarques peuvent permettre d'éviter certains malentendus liés au fondement même de cette controverse. Tout d'abord, nous savons que le fonctionnement des réseaux de neurones biologiques est profondément différent des réseaux de neurones artificiels ; il n'y a pas de confusion possible. Mais cela n'invalide pas l'utilisation du terme « intelligence artificielle ».

Si c'était le cas, il faudrait aussi renoncer à parler de bras mécanique au prétexte qu'un bras biologique est très différent dans son fonctionnement, et que les os, les articulations, les tendons et les muscles ne sont pas des pièces de métal, des poulies, des ressorts et des cordes. Ensuite, on confond souvent l'intelligence et la pensée en général. L'intelligence est une propriété de la pensée, elle n'est pas toute la pensée. L'intelligence a ceci de particulier qu'elle réduit la complexité du monde dans lequel l'être intelligent évolue pour lui permettre de mieux maîtriser son environnement. On se donne des règles pour analyser la réalité, calculer, évaluer et prendre des décisions. Une longue tradition philosophique de penseurs qui ne manquaient pas d'intelligence, n'a cessé de l'affirmer depuis Socrate à Russell en passant par Leibniz. D'une certaine manière, l'intelligence modélise et appauvrit la réalité pour mieux agir dessus, comme une équation en mécanique modélise et appauvrit le mouvement pour mieux le saisir. Enfin, et cela découle de ce qui précède, l'intelligence, même humaine, est dans une large mesure algorithmique : elle analyse des données et calcule selon des procédures. Elle se prête alors très bien à une « mécanisation » et à une « incarnation » au sens littéral du terme : le calcul digital, c'est-à-dire le calcul avec et sur les doigts selon des techniques très variées, est une incarnation du calcul ; avec les différents abaques comme le boulier chinois, la Pascaline[1] et la calculatrice électronique, on assiste à une mécanisation du calcul.

Réfléchir aux buts que nous voulons poursuivre n'est pas seulement une affaire de calcul. Orienter sa vie personnelle et sociale vers certains objectifs qui ont de la valeur, ne relève pas d'une procédure algorithmique. Savoir si nous devrions utiliser des armes nucléaires pour tuer le plus grand nombre de personnes et affaiblir un pays ennemi, cela ne se décide pas uniquement par le calcul des conséquences ; il faut encore définir le bien ou les biens en fonction desquels le calcul des conséquences à un sens moral. Il y a quelque chose de tragique à vouloir éviter la réflexion sur les finalités morales en se contentant d'un calcul sur les moyens. Cette réflexion n'est pas encore à portée d'une intelligence artificielle. Dans le monde que nous connaissons et dans celui que nous pouvons anticiper à court et moyen terme, la réflexion sur les finalités de la vie sociale et de l'existence en général est le produit de l'intelligence humaine.

La *Déclaration de Montréal* pour un développement responsable de l'intelligence artificielle repose sur ce postulat : c'est à l'intelligence humaine et collective de définir les finalités de la vie sociale et en fonction d'elles, les orientations du développement de l'intelligence artificielle afin qu'il soit socialement responsable et moralement acceptable.

[1] Calculatrice mécanique conçue et présentée par le mathématicien et philosophe Blaise Pascal en 1645.

2. POURQUOI LA DÉCLARATION DE MONTRÉAL IA RESPONSABLE ?

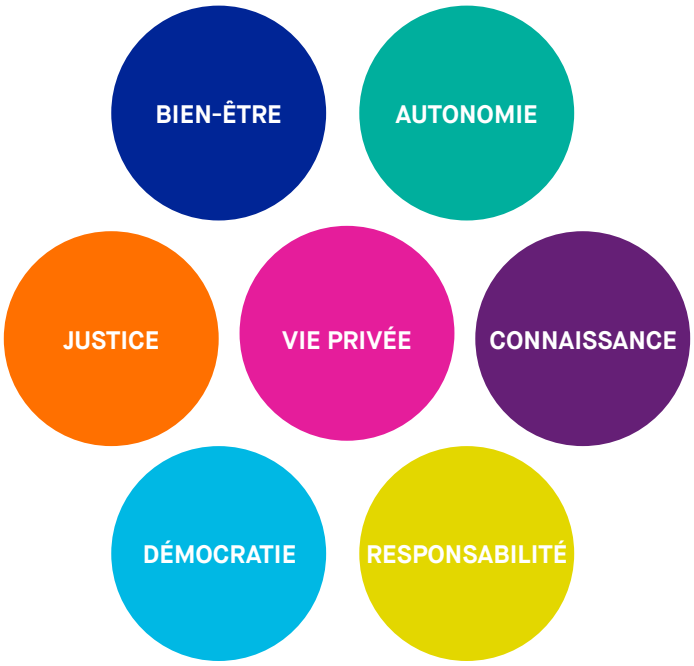
La *Déclaration de Montréal* pour un développement responsable de l'IA est une œuvre collective qui poursuit 3 objectifs :

1. Élaborer un cadre éthique pour le développement et le déploiement de l'IA
2. Orienter la transition numérique afin que tous puissent bénéficier de cette révolution technologique
3. Ouvrir un espace de dialogue national et international pour réussir collectivement un développement inclusif et équitable de l'IA

Il s'agit donc de mettre le développement de l'intelligence artificielle au service du bien-être de tout un chacun, et d'orienter le changement social en élaborant des recommandations reposant sur une légitimité démocratique

La Déclaration est issue d'un processus délibératif inclusif qui met en dialogue citoyens, experts, responsables publics, parties prenantes de l'industrie, des organisations de la société civile et des ordres professionnels. La méthode retenue de la coconstruction citoyenne s'appuie sur une déclaration préliminaire de principes éthiques généraux qui s'articulent autour de valeurs fondamentales.

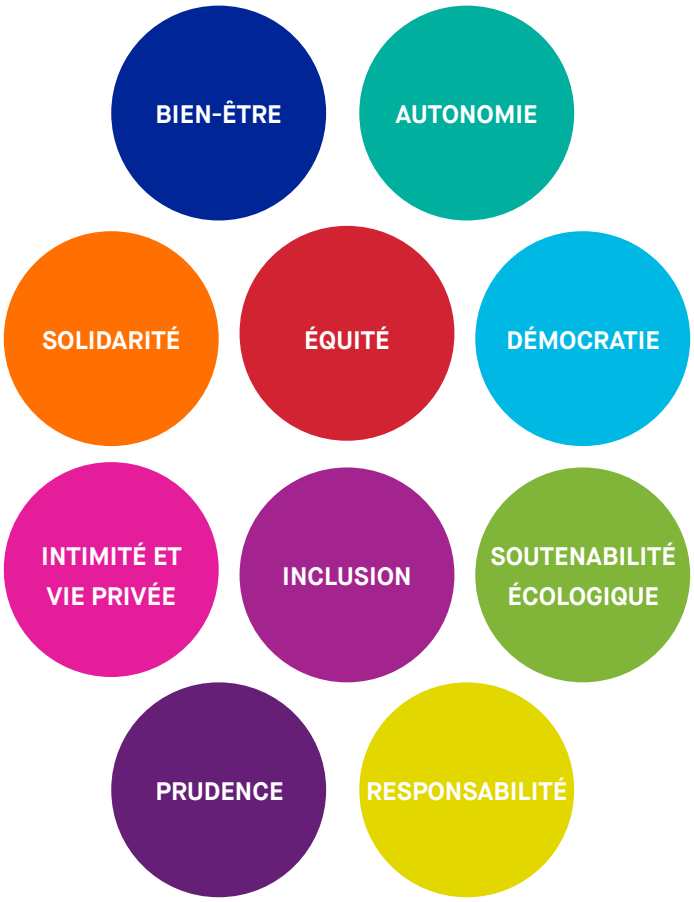
Figure 1: Les valeurs de la Déclaration (version préliminaire)



Notre rapport à ces « valeurs » est ensuite explicité dans des normes qu'on appelle « principes ». Par exemple, si le bien-être est la valeur, notre rapport à cette valeur est celui de la maximisation : nous devons accroître le bien-être des êtres sensibles. Si la valeur est l'autonomie, notre rapport est celui du respect ou de la protection : nous devons respecter l'autonomie des êtres moraux. Le travail initial d'identification de ces valeurs et de ces principes avait pour objectif de lancer un processus de participation citoyenne qui devait alors préciser les principes éthiques d'un développement responsable de l'IA et les recommandations à mettre en œuvre pour s'assurer que l'IA promeuve les intérêts humains fondamentaux.

Au terme de ce processus, la carte des valeurs et des principes a été affinée et permet de se repérer de manière plus précise :

Figure 2: Les valeurs de la Déclaration de Montréal IA responsable



Le passage de la version préliminaire à la version finale de la Déclaration est un travail de réflexion qui s'appuie sur les résultats de la consultation publique et des ateliers de coconstruction. Le choix des valeurs et des principes repose sur une compréhension des attentes sociales fondamentales telles qu'elles ont été exprimées, et il est motivé par le souci de couvrir des enjeux prioritaires, de trouver un équilibre entre les différentes valeurs et par le souci de la cohérence. Parce qu'il n'existe pas de formule toute faite pour parvenir à la sélection des principes (il n'existe pas d'algorithme pour cette tâche), elle est le résultat d'un processus complexe d'ajustement qu'on appelle d'un terme général : délibération.

2.1

L'ORIGINE INTELLECTUELLE DU PROJET

La révolution de l'intelligence artificielle (IA) et plus particulièrement de l'apprentissage profond (*deep learning*) ouvre des perspectives de développement technologique inédites qui permettront d'améliorer les prises de décision, de réduire certains risques et d'offrir une assistance aux personnes les plus vulnérables. Cette révolution est singulière à plus d'un titre, bien qu'elle présente aussi des défis qui se sont déjà posés depuis la fin du XVIII[e] siècle dans l'histoire récente du développement industriel. On aurait tort d'ignorer la spécificité de cette révolution de l'IA en se réfugiant dans des généralités qui ne nous préparent pas à relever les défis actuels. Certes, les humains sont des êtres doués de grandes capacités techniques – l'histoire humaine est elle-même une histoire des transformations techniques de la nature, et l'intelligence artificielle prolonge la tendance à l'automatisation – mais en y regardant de plus près on s'aperçoit que rien ne ressemble à ce qui se joue aujourd'hui avec l'avènement des technologies de l'intelligence artificielle. Les compétences cognitives que l'on croyait réservées aux humains peuvent désormais être exercées par des algorithmes, des machines dont on doit admettre qu'elles sont, en un certain sens, intelligentes.

Les impacts sociaux de ces nouvelles technologies, par ailleurs très diverses, sont encore mal connus. Ils pourraient s'avérer brutaux si nous ne prenions pas dès à présent le temps d'une réflexion éthique, politique, juridique, sociologique ou encore psychologique sur le type de société et de relations humaines que nous voulons promouvoir ou protéger tout en profitant des bénéfices de ces technologies de l'information et du calcul algorithmique.

[1] Paul Meehl, Clinical versus Statistical Prediction, University of Minnesota, 1954.

L'utilisation d'algorithmes pour prendre des décisions techniques ou administratives n'est pas nouvelle. Si les algorithmes sont connus depuis le Moyen Âge[2], l'essor des algorithmes de décision débute véritablement dans les années 1950, en particulier dans le domaine de la santé : triage des urgences dans les hôpitaux, détection des risques de mort subite du nourrisson, prédiction d'accident cardiaque[3]. Toutes ces techniques algorithmiques, « les procédures », posent déjà un certain nombre d'enjeux éthiques et sociaux : celui de l'acceptabilité sociale de la décision « automatique », celui de la dernière décision (un humain est-il au bout de la chaîne de décision ?), ou encore de la responsabilité en cas d'erreur. Et il est évident que ces enjeux se posent de nouveau avec les dernières innovations algorithmiques.

Qu'est-ce qui est différent alors avec les nouvelles technologies que l'on regroupe sous l'acronyme IA ? D'un point de vue objectif, ce qui change c'est la quantité d'informations qui peuvent être traitées par les ordinateurs (les données massives), la puissance des calculateurs et la complexité des algorithmes apprenants qui, se nourrissant de données massives, peuvent accomplir des tâches perceptives et cognitives permettant la reconnaissance visuelle ou auditive, et la prise de décision dans des contextes définis. En combinant les différentes fonctions (reconnaissance faciale, évaluation d'un comportement, décision), les IA présentent des problèmes éthiques particulièrement importants. D'un point de vue subjectif, ce qui est nouveau c'est la prise de conscience citoyenne, aussi tardive que soudaine, des enjeux de la gouvernance algorithmique, du traitement des données personnelles et des impacts sociaux que certains secteurs professionnels subissent déjà.

Si les progrès de l'IA suscitent l'étonnement, voire la fascination, ils éveillent aussi la peur que le recours aux machines, notamment aux robots, appauvrisse considérablement les relations humaines dans les domaines des soins médicaux, de la prise en charge des personnes âgées, de la représentation juridique ou encore, de l'enseignement. Les réactions face au développement de l'intelligence artificielle peuvent même s'avérer hostiles quand l'IA est mise au service d'un contrôle accru des individus et de la société, une perte d'autonomie et une réduction des libertés publiques. Ainsi l'espoir que l'intelligence artificielle soit porteuse de progrès sociaux porte l'ombre d'une crainte : mise entre de mauvaises mains, l'IA pourrait devenir une arme de domination massive (contrôle de la vie privée, concentration de capitaux, nouvelles discriminations). Nombreuses sont les personnes qui émettent également des doutes sur les buts qui animent les chercheurs, les développeurs, les entrepreneurs et les responsables politiques.

Le développement de l'IA et de ses applications met donc en jeu des valeurs éthiques fondamentales qui peuvent entrer en conflit et engendrer des dilemmes moraux graves ainsi que de profondes controverses sociales et politiques : doit-on privilégier la sécurité publique par l'accroissement des moyens de surveillance intelligente (reconnaissance faciale, anticipation des comportements violents) au détriment des libertés individuelles ? Améliorer objectivement le bien-être des individus, notamment en incitant les personnes à adopter des comportements normalisés par les appareils intelligents (comportements alimentaires, gestion du travail, organisation de la journée), peut-il se faire sans respecter leur autonomie ? L'objectif de performance économique doit-il l'emporter sur la préoccupation pour une répartition équitable des bénéfices du marché de l'IA ?

[2] Des procédures algorithmiques sont connus depuis l'Antiquité en fait, mais contrairement à ce que le « th » de algorithme pourrait laisser croire, le mot ne vient pas du grec ancien mais d'une latinisation du nom du mathématicien ayant vécu à Bagdad au 9e siècle : Muhammad Ibn Musa Al-Khwarizmi. Des traductions latines du manuel d'algèbre de Al-Khwarizmi avaient circulé en Europe occidentale dès le 12e siècle dont le premier est le manuscrit de Cambridge *Dixit Algorizmi*. Le manuscrit original en arabe a été perdu. Par déformation, al-Khuwārizmī est donc devenu algorizmi et algoritmi, puis algorithme. Sur l'histoire de ces textes, voir l'édition de référence d'André Allard, *Muhammad Ibn Musa Al-Khwarizmi, Le calcul Indien (algorismus). Versions latines du XIIe siècle*, Librairie scientifique et technique Albert Blanchard, Paris, 1992.

[3] Paul Meehl, *Clinical versus Statistical Prediction*, University of Minnesota, 1954.

Ces dilemmes ou ces tensions ne peuvent être surmontés par une simple hiérarchisation des valeurs et des intérêts fondamentaux. Pour le dire autrement, il ne s'agit pas de classer a priori les valeurs en ordre d'importance, ni de construire une échelle simple et univoque de valeurs, encore moins d'en privilégier certaines en ignorant les autres (la sécurité aux dépens de la liberté, l'efficacité sans la justice sociale, le bien-être au prix de l'autonomie). On ne peut non plus espérer trouver des solutions uniques et définitives. Il convient plutôt de prendre au sérieux les dilemmes moraux causés par le développement de l'IA et de construire collectivement un cadre éthique, politique et juridique qui nous permette d'y faire face en respectant les différentes valeurs fondamentales auxquelles nous tenons légitimement comme membres d'une société démocratique.

2.2

LE FORUM SUR LE DÉVELOPPEMENT SOCIALEMENT RESPONSABLE DE L'INTELLIGENCE ARTIFICIELLE

Ces réflexions ont été le point de départ de l'initiative des Fonds de recherche du Québec et de l'Université de Montréal pour organiser une rencontre internationale faisant le point sur les impacts sociaux de l'IA. Dans ce cadre, le comité d'organisation à l'Université de Montréal proposait de lancer les travaux de la *Déclaration de Montréal pour un développement responsable de l'IA* sur la base d'un processus de consultation et de participation[4]. Les 2 et 3 novembre 2017, s'est ainsi tenu au Palais des congrès de Montréal, un forum rassemblant les plus grands experts des domaines concernés par la réflexion sur l'IA, des sciences fondamentales aux sciences humaines et sociales.

Le Forum proposait d'établir les balises d'une réflexion collective sur le développement éthique et socialement responsable de l'intelligence artificielle, en poursuivant les trois objectifs suivants :

> Offrir un espace de réflexion public quant aux enjeux du développement de l'IA et ses impacts sociaux

> Intéresser et sensibiliser les décideurs, les partenaires industriels, les représentants politiques et la communauté qui s'intéressent à l'IA quant aux questions de société soulevées par son essor et ses applications

> Valoriser une approche interdisciplinaire et intersectorielle comme facteur de réussite essentiel au développement éthique et durable de l'IA

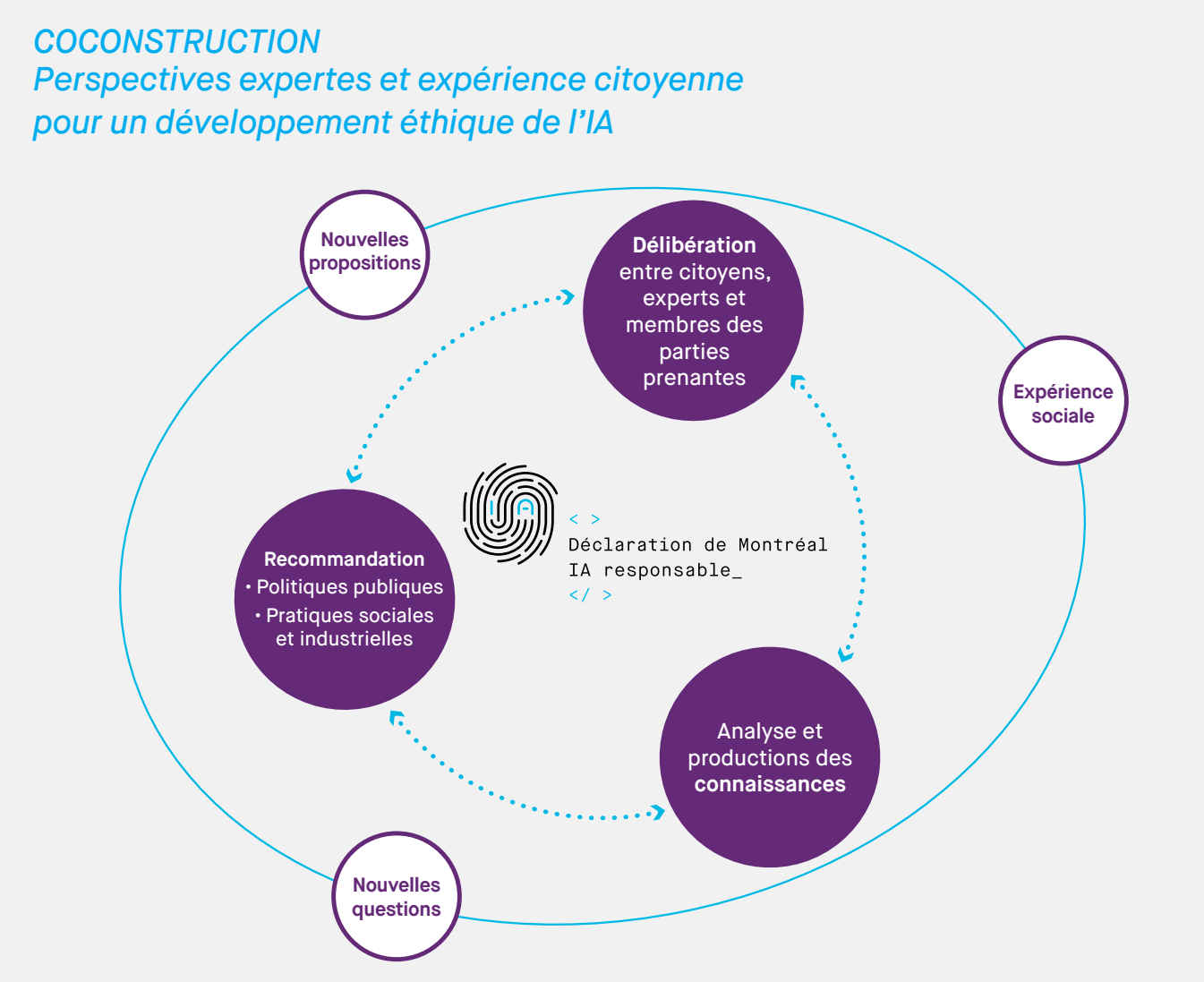
Se sont ainsi définis les contours d'une démarche inclusive (interdisciplinaire et intersectorielle) qui est au cœur de l'entreprise d'élaboration de la *Déclaration de Montréal* pour un développement de l'IA qui soit à la fois responsable, vecteur de progrès social et garant de l'égalité et de la justice. La version préliminaire de cette *Déclaration de Montréal* fut présentée en clôture du Forum. Il s'agissait alors de lancer un processus de coconstruction citoyenne autour de l'éthique de l'IA, processus que nous détaillerons dans la section 3.

[4] Le comité scientifique du Forum était composé de Louise Béliveau (UdeM, Vice-rectorat aux affaires étudiantes et aux études), Yoshua Bengio (UdeM, Département d'informatique, MILA, IVADO), David Décary-Hétu (UdeM, École de criminologie), Nathalie De Marcellis-Warin (École Polytechnique, Département de mathématiques et de génie industriel, CIRANO – Centre interuniversitaire de recherche en analyse des organisation), Marc-Antoine Dilhac (UdeM, Département de philosophie, CRÉ Centre de recherche en éthique), Marie-Josée Hébert (UdeM, Vice-rectorat à la recherche, à la découverte, à la création et à l'innovation), Gregor Murray (UdeM, École de relations industrielles et CRIMT – Centre de recherche interuniversitaire sur la mondialisation et le travail), Doina Precup (Université McGill, School of Computer Science; MILA), Catherine Régis (UdeM, Faculté de droit, CRDP – Centre de recherche en droit public), Christine Tappolet (UdeM, Département de philosophie et CRÉ – Centre de recherche en éthique).

2.3

VERS LA DÉCLARATION DE MONTRÉAL POUR UN DÉVELOPPEMENT RESPONSABLE DE L'IA

Figure 3 : La démarche de coconstruction



Comme nous l'avons mentionné en début de chapitre, le travail initial d'identification de ces valeurs et des principes correspondants avait pour seul objectif de lancer un processus de participation citoyenne afin de préciser les principes éthiques d'un développement responsable de l'IA, de les enrichir et de les compléter. On ne s'étonnera donc pas que la version préliminaire de la Déclaration soit schématique et que l'énoncé des principes soit volontairement très simple et consensuel, laissant à la délibération publique la possibilité de les interpréter et de les compléter[5]. Un an plus tard, la Déclaration a été substantiellement enrichie.

[5] Le comité scientifique en charge de la rédaction de cette version préliminaire était composé de Yoshua Bengio (UdeM, Département d'informatique, MILA, IVADO), Guillaume Chicoisne (IVADO), Marc-Antoine Dilhac (UdeM, Département de philosophie, CRÉ Centre de recherche en éthique), Vincent Gautrais (UdeM, Faculté de droit, CRDP – Centre de recherche en droit public), Martin Gibert (CRÉ – Centre de recherche en éthique, IVADO), Pascale Lehoux (UdeM, ESPUM – École de santé publique), Joëlle Pineau (Université McGill, School of Computer Science; MILA), Peter Railton (Université du Michigan, Académie américaine des arts et des sciences, philosophie), Christine Tappolet (UdeM, Département de philosophie et CRÉ – Centre de recherche en éthique).

Si l'un des objectifs du processus de coconstruction était d'affiner les principes éthiques proposés dans la version préliminaire de la *Déclaration de Montréal*, un autre objectif tout aussi important consistait à élaborer des recommandations pour encadrer la recherche en IA et son développement technologique et industriel. Cependant, il est trop fréquent de voir les rapports d'analyse et de recommandations oubliés aussitôt qu'ils sont publiés : il est donc crucial de ne pas perdre l'élan public manifesté au cours de la période de coconstruction.

Dès lors que le processus de coconstruction est achevé (ou suspendu), il est nécessaire que s'ouvre un débat public dans les lieux où les décisions politiques, juridiques et réglementaires sont prises, afin d'approfondir les analyses et de mettre en œuvre concrètement les pistes de solution et les recommandations issues de la délibération citoyenne. Ces recommandations ne sont pas uniquement de nature juridique et, quand elles le sont, elles n'impliquent pas nécessairement une modification de la loi. Elles peuvent cependant demander une modification du cadre légal ; dans certains domaines, elles le doivent. Dans d'autres cas, les recommandations auront pour objectif de nourrir et d'orienter la réflexion des organisations professionnelles afin qu'elles modifient leur code de déontologie, ou que les entreprises adoptent un nouveau cadre éthique.

Cette étape est donc le but final du processus de coconstruction. Il faut toutefois préciser que face à une technologie qui n'a cessé de progresser depuis 70 ans et dont les innovations majeures se succèdent à présent tous les 2 à 5 ans en moyenne, il serait déraisonnable de présenter la Déclaration comme définitive et complète. Il est essentiel de penser la coconstruction comme un processus ouvert, avec des phases successives et cycliques de délibération, de participation et de production de recommandations, et de concevoir la Déclaration elle-même comme un document d'orientation révisable et adaptable en fonction de l'évolution des connaissances et des techniques de l'intelligence artificielle. Ce processus de production de connaissances, de délibération citoyenne et de recommandations d'encadrement éthique et de politique publique, devra se prolonger dans une structure institutionnelle pérenne qui permette de rester réactif face aux évolutions de l'IA.

2.4

MONTRÉAL ET LE CONTEXTE INTERNATIONAL

L'initiative de la *Déclaration de Montréal* s'inscrit dans un contexte scientifique, social et industriel favorable. Montréal est un pôle de recherche et de développement majeur en intelligence artificielle avec une communauté de chercheurs et des laboratoires universitaires de réputation mondiale (MILA, IVADO) et une pépinière de *start-ups* et d'entreprises en plein essor. Ce développement scientifique, technologique et industriel est au cœur d'une révolution des pratiques sociales, des modèles économiques et des modes de vie, qui touche tous les secteurs de la société. La Ville de Montréal, avec son Laboratoire de l'innovation urbaine[6], est aussi un laboratoire vivant du changement social et technologique. Avec la recherche scientifique fondamentale viennent des responsabilités éthiques et sociales que la communauté montréalaise de l'IA assume pleinement.

Mais au-delà de Montréal, c'est tout le Québec et le Canada qui offrent le contexte social propice pour s'engager dans une réflexion sur les impacts sociaux de l'IA. Comme le MILA à Montréal, Vector à Toronto, AMII (Alberta Machine Intelligence Institute) à Edmonton, et le CRDM (Centre de recherche en données massives) à Québec constituent des pôles d'excellence dans la recherche fondamentale qui ont entraîné une croissance industrielle extrêmement rapide et robuste. L'Institut canadien de recherches avancées (ICRA|CIFAR), partenaire des travaux de la Déclaration, a joué un rôle de première importance dans ce développement canadien de l'IA en soutenant la recherche fondamentale quand l'IA traversait son « hiver ». L'initiative de la Déclaration est ainsi portée par différents acteurs québécois et canadiens hors Montréal.

[6] http://ville.montreal.qc.ca/portal/page?_pageid=5798,141982209&_dad=portal&_schema=PORTAL

Plusieurs interlocuteurs internationaux ont également manifesté leur intérêt pour la *Déclaration de Montréal*, notamment pour sa méthode d'élaboration. L'équipe de la Déclaration a pu établir un dialogue avec des institutions comme la *Royal Society* du Royaume-Uni[7], l'EGE (*European Group on Ethics in Science and New Technologies*[8]) et le HLEG (*High Level Expert Group on AI*[9]) de la Commission européenne qui ont leur propre programme d'étude et de recommandation sur l'IA. On note d'abord une convergence des lignes directrices pour un développement éthique de l'IA ainsi qu'une volonté partagée de faire valoir une conception démocratique de l'utilisation de l'IA au service du bien commun.

La démarche de la *Déclaration de Montréal* doit ainsi se comprendre dans le contexte international d'un **printemps de l'IA**. Elle est précédée par plusieurs initiatives qui doivent être saluées car elles ont catalysé la réflexion sur une IA responsable. Il faut tout d'abord évoquer la création en 2014 du *Future of Life Institute* qui a produit en 2017 la Déclaration d'Asilomar : à l'issue d'une conférence de 3 jours, une déclaration contenant 23 principes fondamentaux encadrant la recherche sur l'IA et ses applications a été signée par plus de 1200 chercheurs. Y participait alors le professeur Yoshua Bengio qui attirait l'attention sur les risques d'utilisation irresponsable et malveillante de l'IA[10].

Depuis la conférence d'Asilomar, plusieurs rapports sur l'éthique de l'IA ont été publiés. Le rapport de l'Association internationale des ingénieurs électriciens et électroniciens (IEEE), *Ethically aligned design*. V2, a été rendu public fin 2017 et a réuni plusieurs centaines d'ingénieurs et de chercheurs en IA. L'Institut *AI Now* basé à la New York University a également produit plusieurs rapports, dont le dernier porte sur l'évaluation des impacts de l'IA[11]. Deux rapports stratégiques ambitieux ont été publiés en mars et avril 2018 : le rapport de la Mission Villani en France et celui de la Chambre des lords au Royaume-Uni « AI in the UK : ready, willing, and able ? ». Il faut aussi souligner la démarche participative de la CNIL (Commission nationale de l'informatique et des libertés) en France qui a débouché sur la publication du rapport au titre évocateur : « Comment permettre à l'Homme de garder la main ? – Les enjeux éthiques des algorithmes et de l'intelligence artificielle », en décembre 2017.

Comment se positionne la *Déclaration de Montréal* dans ce concert d'initiatives indépendantes ? Et que penser de l'inflation éthique autour de l'IA ? Cette dernière question est d'autant plus importante que nous partageons la mise en garde de l'EGE dans son rapport *Artificial Intelligence, Robotics and 'Autonomous' Systems* (mars 2018) qui rappelle qu'en l'absence d'une réflexion coordonnée sur les enjeux éthiques et sociaux de l'IA, il existe un risque de « ethics shopping »[12]. La conséquence immédiate serait une forme de délocalisation des coûts éthiques dans les régions du monde où les critères éthiques sont les moins exigeants. Un autre risque est aussi une forme de banalisation du discours éthique.

Chaque processus d'élaboration d'un cadre éthique a ses mérites. La partie 2 de ce rapport dresse un « Portrait 2018 des recommandations internationales en éthique de l'IA ». La particularité de la démarche de la *Déclaration de Montréal* est d'être

[7] Nous tenons à remercier Natasha McCarthy (Head of Policy) et Jessica Montgomery (Senior Policy Adviser) de la Royal Society d'avoir permis ce dialogue, ainsi que le UK Science and Innovation Network in Canada qui l'a facilité.

[8] Le European Group on Ethics in Science and New Technologies (EGE) est un organe indépendant de réflexion et de conseil pour le Président de la Commission européenne. Nous remercions la Délégation générale du Québec à Bruxelles ainsi que Mission Canada auprès de l'UE d'avoir rendu possible plusieurs rencontres entre juin et novembre 2018.

[9] Le HLEG on AI est un groupe de 52 experts retenus par la Commission européenne pour définir les principes d'application de la stratégie européenne de l'IA. Nous remercions les responsables du HLEG de nous avoir permis de participer à leurs travaux entre septembre et novembre 2018, afin de partager et d'enrichir nos réflexions et nos expériences respectives.

[10] Entrevue de Yoshua Bengio lors de la conférence d'Asilomar : futureoflife.org/2017/01/18/yoshua-bengio-interview/

[11] AI Now Institute, Algorithmic Impact Assessments : A Practical Framework For Public Agency Accountability, Avril 2018.

[12] EGE, Artificial Intelligence, Robotics and 'Autonomous' Systems (mars 2018), p. 14.

essentiellement participative. De février à novembre 2018, le processus de coconstruction a mobilisé, au Québec et en Europe, plus de 500 citoyens, experts et parties prenantes au cours d'une quinzaine d'ateliers, de journées de coconstruction et de tables rondes. Si d'autres initiatives de type participatif ont été menées ailleurs, en particulier en France, celle de la *Déclaration de Montréal* se distingue par son ampleur et par ses méthodes prospectives.

La *Déclaration de Montréal* a pour vocation d'ouvrir un espace de dialogue au Québec et au Canada et d'offrir, au-delà des frontières canadiennes, une plateforme de réflexion commune. L'objectif est de dégager les orientations socialement acceptables et innovantes de l'IA en prenant pour point de départ la réflexion citoyenne informée dans les différentes démocraties concernées. Il faut aussi que cet espace de dialogue soit accessible aux citoyens des sociétés moins démocratiques qui manifestent leur désir de participer à un débat global sur le futur des sociétés humaines.

3. LES ENJEUX ÉTHIQUES ET SOCIÉTAUX DE L'IA

Le processus de réflexion collective au cœur de l'élaboration de la Déclaration de Montréal s'appuie sur la version préliminaire de la Déclaration de principes éthiques elle-même et sur des exposés informatifs sur l'IA et l'éthique de l'IA.

3.1

SE FAIRE UNE IDÉE DE L'IA

L'idée de l'IA n'est pas nouvelle. Il faudrait au moins remonter au 17e siècle et à l'idée d'une caractéristique universelle et d'un art combinatoire du philosophe et mathématicien Leibniz : raisonner revient à calculer et la pensée est conçue de manière algorithmique[13]. La notion de *calculus ratiocinator* (le calcul logique) préfigure l'idée de machine intelligente telle qu'elle sera développée trois siècles plus tard, dans les années 1940, par Alan Turing. En 1948, dans un rapport intitulé « *Intelligent Machinery* » et en 1950, dans son fameux article « *Computing Machinery and Intelligence*[14] » , Alan Turing évoque l'intelligence de la machine et élabore le jeu de l'imitation pour définir les conditions dans lesquelles on peut dire qu'une machine pense. Le terme d'intelligence artificielle apparait pour la première fois en 1955 dans le descriptif d'un atelier de travail proposé par John McCarthy (Dartmouth College), « 2 months, 10 man study of artificial intelligence ». Mais les applications et les possibilités de développement de l'IA semblent alors très limitées, et l'*hiver de l'IA* débute, avec un intérêt moindre de la part de la communauté scientifique. Pourtant, si le développement de la discipline reste timide en comparaison de l'effervescence philosophique et culturelle qu'elle suscite (pensons à *2001 : A Space Odyssey, Blade Runner* ou *Terminator* pour ne citer que des films populaires), les recherches n'ont jamais cessé dans ce domaine et il faut attendre le début du 21e siècle pour assister au printemps de l'IA.

L'IA consiste d'une certaine manière à simuler l'intelligence humaine[15], s'en inspirer et la reproduire. Mais dans un premier temps, c'est le cerveau, siège de l'intelligence humaine, qui a été conçu comme une machine capable de recueillir, percevoir, et collecter des données de son environnement qu'il va ensuite analyser, interpréter et comprendre, se nourrissant de ces expériences pour établir des relations. Le domaine de recherche de l'IA consiste à produire des outils mathématiques pour formaliser les opérations de l'esprit et ainsi créer des machines qui peuvent accomplir des tâches cognitives plus ou moins générales, associées à l'intelligence humaine naturelle. Par exemple, découvrir des motifs complexes parmi une grande quantité de données, ou encore raisonner de manière probabiliste, afin de classer en fonction de catégories des informations, de prédire une donnée quantitative ou de regrouper des données ensemble. Ces compétences cognitives sont à la base d'autres compétences comme celles de décider entre plusieurs actions possibles pour réaliser un objectif, d'interpréter une image ou un son, de prédire un comportement, d'anticiper un événement, de diagnostiquer une pathologie, etc.

Mais ces compétences cognitives ne sont possibles que si la machine est aussi capable de percevoir des formes sensibles comme les images et les sons, ce qui est rendu possible par les récentes innovations informatiques. La notion d'IA couvre donc aussi les technologies de reconnaissance visuelle ou auditive

[13] Leibniz (1666), *De Arte combinatoria* (« De l'art combinatoire »).

[14] A. M. Turing (1950), « Computing Machinery and Intelligence ». *Mind* 49, p. 433-460.

[15] Alan Turing ouvre ainsi son rapport « Intelligent Machinery » (1948) : « I propose to investigate the question as to whether it is possible for machinery to show intelligent behaviour. »

qui permettent à la machine de percevoir son environnement et d'élaborer une représentation de cet environnement.

Ces réalisations de l'IA reposent sur deux éléments : des données et des algorithmes, c'est-à-dire des suites d'instructions permettant d'accomplir une action complexe. Pour schématiser, si vous voulez cuisiner un nouveau plat, il vous faut connaître les ingrédients (les données) et suivre une recette qui donne des instructions pour les utiliser correctement (l'algorithme). Jusqu'à présent, les capacités de traitement des données (quantité de données et algorithmes de traitement) étaient trop limitées pour envisager un développement utile des technologies de l'IA. Les choses ont changé avec l'utilisation de matériaux rendant possible la construction de calculateurs très petits et très rapides (les puces électroniques) et le stockage d'immenses quantités de données, et avec l'avènement d'une ère de l'information grâce à internet.

Ce qui a changé, c'est la quantité gigantesque de données que l'on est en mesure de générer, de transmettre, mais aussi de traiter. Si les données massives (*big data*) existaient déjà dans le passé, par exemple dans l'industrie financière, aujourd'hui c'est une multitude d'objets inanimés, de lieux, ou de capteurs qui produisent en tout temps des données structurées ou non, qu'il faut manipuler et transformer avant de pouvoir les exploiter (*data mining*). Il peut s'agir de l'observation de millions de messages publiés sur les réseaux sociaux, de l'ensemble des mots provenant d'une bibliothèque de milliers d'œuvres, ou encore du contenu d'énormes banques d'images.

Mais ce qui a changé, c'est aussi le type d'algorithme élaboré par les chercheurs en IA. Les algorithmes déterministes, qui sont une suite déterminée d'instructions comme une recette de cuisine, laissent désormais la place à des algorithmes apprenants et reposent sur des réseaux neuronaux de plus en plus complexes à mesure que la puissance de calcul des machines augmente. En informatique, on parle de *machine learning* (apprentissage machine ou apprentissage automatique) et les progrès de ce secteur de recherche ont été renforcés par le développement du *deep learning* (apprentissage profond). Au cœur de la notion même d'IA, se trouve la capacité d'adaptation et d'apprentissage. En effet, pour qu'une machine puisse être considérée comme intelligente, il faut qu'elle soit capable d'apprendre par elle-même à partir des données qui la nourrissent, comme le fait un être humain. Et comme pour l'être humain, l'apprentissage machine peut être supervisé, ou non supervisé, par des êtres humains qui entraînent les machines sur les données.

Ce sont ces techniques de *deep learning* qui ont permis aux machines de surpasser les êtres humains dans des jeux complexes comme les échecs avec AlphaZero, qui bat d'ailleurs toutes les autres machines qui n'utilisent pas le *deep learning*, et le jeu de Go qui était réputé indomptable pour les algorithmes, mais qui a vu le triomphe de AlphaGo sur les meilleurs joueurs mondiaux à partir de 2015.

Si ces exemples sont éloquents, l'usage de l'IA sert d'autres buts comme l'automatisation de tâches qui nécessitent jusqu'à présent l'intervention humaine, en particulier des tâches de perception et de reconnaissance. Par exemple : le traitement de la parole, la reconnaissance d'objets, de mots, de formes, de texte, l'interprétation des scènes représentées, des couleurs, des similarités ou des différences dans de grands ensembles, et par extension l'analyse de données et la prise de décision – ou l'aide à la prise de décision. Les possibilités sont très vastes, et sont décuplées à mesure que les ingénieurs et les informaticiens les combinent pour créer de nouvelles utilisations.

3.2

L'IA AU QUOTIDIEN ET LE QUESTIONNEMENT PHILOSOPHIQUE

L'IA nous engage dans une réflexion éthique qui, à la différence de celle sur le nucléaire ou la génomique, porte sur des objets et des technologies du quotidien. L'IA est devenue omniprésente et façonne plus que jamais nos vies. Nous sommes habitués à porter de petits objets connectés (téléphones, montres) et nous nous préparons à l'arrivée de véhicules autonomes, voitures et bus, mais nous prenons déjà des trains et des métros qui fonctionnent de manière autonome, et les avions sont capables, en pilotage automatique, de décoller, naviguer et atterrir sans intervention humaine. Nous utilisons des algorithmes de classement pour nos recherches sur internet, des correcteurs orthographiques intégrés à nos messageries, des applications de recommandation pour la musique ou les rencontres, et nous savons que les administrations utilisent des algorithmes de triage, les banques des algorithmes de gestion et placements financiers et que certains diagnostics médicaux peuvent désormais être réalisés avec une grande précision par les algorithmes, etc.

Ces technologies sont si bien intégrées dans notre quotidien que nous n'y pensons plus vraiment. Quand on évoque l'IA, la plupart de gens l'associent encore à des machines menaçantes, polyvalentes et dotées d'une forme de conscience, capables de former un plan pour se débarrasser des êtres humains[16].
Or l'expérience de l'IA est tout à fait banale aujourd'hui, les algorithmes de recommandations envahissent internet (Google, Amazon, Facebook).
Si vous magasinez en ligne sur internet, il y a de fortes chances qu'une fenêtre d'aide s'ouvre et que Inès commence la conversation par :

« Boujour, que puis-je faire pour vous aider ? ».

« Bonjour Inès »

Pendant quelques instants, vous avez l'impression qu'une personne, du nom d'Inès, vous parle derrière son écran ; pendant quelques instants, le doute est permis. Inès vous pose des questions, répond aux vôtres, vous fournit les informations importantes dont vous avez besoin pour faire votre magasinage. Mais après quelques échanges, on se rend compte que si Inès livre les informations pertinentes disponibles, elle semble répondre de façon mécanique, elle ne comprend pas la manière dont vous parlez, elle ne saisit pas l'humour ou les questions décalées, en d'autres termes elle n'interagit pas vraiment avec vous de manière naturelle. Inès est une agente conversationnelle, un *chatbot*, une IA. Il est devenu banal de discuter en ligne avec des *chatbot*s pour demander des informations sur son assurance maladie ou un nouveau plan bancaire, ou encore pour demander un conseil vestimentaire.

Pour l'instant, les *chatbot*s sont repérables après quelques minutes de conversation, souvent moins. Si un *chatbot* réussissait à ne pas être détecté par un humain pendant un temps raisonnable, nous pourrions considérer que cette machine a passé avec succès le test de Turing et nous aurions alors, selon ce test, un cas d'intelligence artificielle, c'est-à-dire de machine qui pense.

Dans son célèbre article, « Computing Machinery and Intelligence », le père de l'informatique moderne, Alan Turing, se propose de répondre à la question : « Une machine peut-elle penser ? »[17]. Or, dès l'introduction de son article, il change le problème auquel il estime pouvoir donner une solution : une machine peut-elle se comporter de telle sorte qu'on ne puisse pas faire la différence avec une personne humaine ? Il propose alors le fameux « jeu de l'imitation » qui consiste à mettre en communication un être humain qui pose des questions (l'interrogateur) avec un autre être humain et une machine qui répondent à ses questions. Si la machine imite assez bien l'être humain au point que l'interrogateur ne parvient pas à dire qui de l'être humain ou de la machine a répondu, nous pouvons considérer que la machine pense. C'est cela qu'on désigne par l'expression « test de Turing ».

[16] Stanley Kubrick a magistralement capté (et contribué à former) cet imaginaire avec le très humain ordinateur HAL 9000, dans son film *2001 : A Space Odyssey* (1968).

[17] A. M. Turing (1950).

Ce jeu de l'imitation a fait couler beaucoup d'encre et les philosophes se sont durement opposés les uns aux autres pour savoir si nous pouvions dire qu'une machine pense. Une expérience connue sous le nom de « la chambre chinoise » a été popularisée dans les années 1980 par le philosophe John Searle[18]. Selon Searle, une machine qui agit extérieurement de la même façon qu'un être humain ne peut être considérée comme possédant une intelligence au sens fort du terme. Imiter un comportement intentionnel n'est pas la même chose qu'agir de manière intentionnelle. Pour illustrer ce point, Searle nous demande d'imaginer une chambre dans laquelle se trouve une personne qui, ne connaissant rien du chinois, va se faire passer pour un locuteur chinois. C'est une variante du jeu de l'imitation : la personne dans la chambre chinoise, appelons-le John, reçoit des messages écrits en chinois que des locuteurs chinois à l'extérieur de la chambre lui transmettent. John ne comprend rien aux messages qu'il reçoit, mais possède un manuel d'instruction très complet qui lui permet de manipuler les signes chinois et de composer des réponses qui sont comprises par les locuteurs chinois à l'extérieur de la chambre, de sorte que ces derniers pensent que les réponses ont été écrites par une personne comprenant le chinois. Searle en conclut que dans ce cas John a simulé la compétence linguistique mais qu'il ne la possède pas ; il a fait croire qu'il comprenait le chinois, mais il ne comprenait pas ce qu'il écrivait. Il faut, selon Searle, appliquer la même conclusion pour l'IA : une machine intelligente manipule des signes, elle suit un algorithme, c'est-à-dire une suite d'instructions pour accomplir une tâche (ici parler), mais elle ne comprend pas ce qu'elle fait.

Ce débat est fascinant et il est loin d'être réglé, mais on n'a pas vraiment besoin de trancher la question que posait Turing pour s'interroger sur la place de l'IA dans nos vies et dans nos sociétés. Pour l'instant, les *chatbot*s bien entraînés font aussi bien que les êtres humains dans un cadre de conversations très limitées, mais elles ne font pas illusion quand ce cadre change. Et même si l'IA inaugure une ère où il est de plus en plus difficile de distinguer un être intelligent naturel d'un être artificiellement intelligent, les machines intelligentes restent des outils développés pour accomplir des tâches bien définies. On peut donc laisser à la philosophie cognitive, à la métaphysique, la psychologie et aux neurosciences le soin de débattre de la notion d'intelligence artificielle et discuter de la possibilité que les robots développent des émotions et éprouvent de l'empathie[19]. Le questionnement que pose l'introduction des IA dans nos vies est de type pratique, qu'il soit éthique, politique ou juridique. C'est un questionnement d'abord sur les valeurs et les principes éthiques, sur les orientations des politiques publiques et sur l'application de normes pour encadrer la recherche en IA et ses applications.

Mais parce que les technologies de l'IA sont indifférentes à leurs multiples applications, le problème n'est pas de savoir si l'IA est bonne ou mauvaise en soi, mais de déterminer quels usages et quels objectifs sont éthiques, socialement responsables, compatibles avec les valeurs et les principes politiques démocratiques. Cependant, cette réflexion éthique ne concerne pas seulement les applications de l'IA, elle porte aussi sur la recherche en IA, ses orientations générales et ses buts. La recherche sur le nucléaire n'était pas initialement destinée à produire des bombes d'une puissance tragique pour l'humanité. Mais plusieurs programmes scientifiques avaient ce but. Il faut donc être particulièrement attentif à la direction que prend la recherche en IA, celle qui se fait à l'université comme celle qui est développée par les entreprises privées ou par des organismes gouvernementaux.

[18] J. Searle (1980), 'Minds, Brains and Programs'. *Behavioral and Brain Sciences 3*, p. 417–57.

[19] Ce qui est très différent des questions sur l'usage des machines pour détecter les émotions humaines, les traiter et y répondre de manière adéquate. Voir par exemple les travaux de Rosalind W. Picard, *Affective Computing*, Cambridge, MIT Press, 1997.

3.3

LES ENJEUX ÉTHIQUES DE L'IA

Pourquoi introduire l'éthique quand on aborde les impacts sociétaux, sociaux et économiques de l'IA ? Peut-on se payer le luxe d'une réflexion éthique ? Et n'est-il pas un peu naïf de vouloir encadrer avec des principes éthiques le développement de l'IA qui génère des profits colossaux ? Ces questions, les éthiciens les entendent fréquemment parmi des citoyens dubitatifs et aussi parmi des décideurs qui font l'expérience des limites de leur marge de manœuvre. Pour y répondre, il faut d'abord présenter très brièvement le domaine de l'éthique quand on aborde les enjeux sociétaux de l'IA.

Pour faire simple, l'éthique est une réflexion sur les valeurs et les principes qui sous-tendent nos décisions et nos actions, quand elles affectent les intérêts légitimes d'autrui. Cela suppose que l'on puisse s'entendre sur les intérêts légitimes des personnes et c'est précisément ce qui nourrit le débat en éthique. Le domaine de l'éthique ne porte donc pas sur ce que l'on peut faire, mais en général, sur ce que l'on doit faire, ou devrait faire : on peut tuer un million de personnes avec une seule bombe nucléaire, mais doit-on le faire pour impressionner un pays ennemi et démoraliser une population en guerre ? Prenons un exemple moins tragique : on peut mentir à un ami au sujet de sa nouvelle coupe de cheveux, mais est-il moral de lui mentir pour lui épargner une déception ? Que doit-on faire dans ce cas ? Pour répondre à cette question, il faut examiner les options que l'on a : dire la vérité, ou ne pas la dire, ou ne pas dire toute la vérité, ou encore la dire d'une certaine manière. Il faut examiner les conséquences aussi de chaque option, se demander si elles sont importantes, pourquoi elles le sont. Il faut aussi réfléchir aux objectifs qui ont de la valeur (faire du bien à autrui, respecter autrui). Il faut enfin se donner une règle, un principe moral : par exemple, le principe catégorique selon lequel il est toujours mal de mentir, peu importe les conséquences ; ou bien le principe hypothétique selon lequel il n'est pas moralement correct de mentir sauf si…

Le domaine de l'éthique qui s'applique aux enjeux de l'IA est celui de l'éthique publique. Si on recourt au même type de réflexion en éthique publique, l'objet n'est pas le même et le contexte de réflexion non plus : l'éthique publique concerne toutes les questions qui impliquent des choix collectifs difficiles sur des pratiques sociales et institutionnelles controversées qui concernent tous les individus en tant que membre de la société, et non d'un groupe particulier : un médecin doit-il dire à son patient la vérité sur son état de santé même si cela a pour conséquence de le déprimer et d'accélérer la maladie ? Cette question ne porte pas sur la moralité privée du médecin, mais sur le type de comportement et d'acte que l'on est en droit d'attendre d'une personne qui occupe la fonction sociale de médecin. Cette question est de nature publique et devrait faire l'objet d'un débat public pour délimiter à partir des valeurs sociales les bonnes pratiques en matière de relation patient-médecin. Par débat public, on entend toute discussion qui peut prendre des formes diverses de consultation, délibération ou participation démocratique, et qui est ouverte à une diversité d'acteurs individuels et institutionnels comme des professionnels du milieu de pratique, des représentants associatifs ou syndicaux, des experts, des décideurs publics, des citoyens. L'éthique publique appelle à une réflexion collective pour dégager les principes des bonnes pratiques et exige que les acteurs justifient leurs propositions sur la base d'arguments acceptables dans un contexte de pluralisme. Dans le cas du mensonge médical, on peut faire appel à des valeurs partagées comme celle d'autonomie, de respect des personnes, de dignité, de bien-être ou de santé du patient, etc. À partir de ces valeurs, il est possible de construire des principes qui encadrent la pratique médicale et donnent des pistes de régulation par la mise en œuvre d'un code de déontologie, par une modification de la loi ou la promulgation d'une nouvelle législation.

L'éthique publique n'est pas à côté ni au-dessus du droit qui a sa propre logique, mais elle permet de clarifier des enjeux de la vie sociale que différents acteurs doivent avoir à l'esprit pour répondre aux

attentes normatives des citoyens et assurer une coopération sociale équitable. En ce sens, l'éthique publique façonne les politiques publiques, et peut se traduire dans une législation, une règlementation, un code de déontologie, un mécanisme d'audit, etc. Dans le domaine de l'IA, c'est ce type de réflexion éthique que nous mettons en œuvre. Prenons l'exemple de Melody, une agente conversationnelle médicale. Melody fait des diagnostics en ligne, accessible sur votre téléphone cellulaire, en fonction des symptômes que vous lui décrivez. D'une certaine manière elle agit comme un médecin. Cela peut être très pratique dans une société où le système de santé public est peu développé ou peu accessible. Mais que cela soit pratique n'est pas suffisant pour autoriser la mise en service sur le marché d'une application comme Melody. En effet, cette application pose des questions éthiques qu'on ne se posait pas immédiatement avec Inès, le *chatbot* de conseil de vente. Par exemple, on devrait se demander si Melody doit donner à son utilisateur les différents diagnostics possibles, même si celui-ci n'est pas en mesure de comprendre l'information. Ce problème est une simple transposition d'un questionnement en éthique médicale qui a déjà reçu une réponse normative pour laquelle il y a un large consensus. La notion de décision informée, de décision libre et éclairée du patient a permis de préciser les devoirs du médecin. Cela résout-il le problème que posent Melody et ses applications jumelles qui se multiplient de manière souvent peu contrôlée[20]? Dans les grandes lignes, sans doute, mais une attention particulière à cette technologie montre que ce n'est pas aussi simple. Le contexte ne permet pas à Melody de s'assurer que le patient comprenne le diagnostic, ni l'urgence ou non de traiter la pathologie diagnostiquée. Quelles règles inventées pour garantir le bien-être et l'autonomie du patient ? C'est tout l'enjeu de la délibération collective sur les enjeux éthiques de l'IA.

D'autres enjeux sont spécifiques à l'IA et n'ont pas encore reçu de solutions éthiques. Par exemple, si Melody se trompe sur le diagnostic et que l'état de santé de l'utilisateur qui a suivi ses recommandations se dégrade gravement, qui est responsable ? Dans le cas d'une consultation médicale avec un médecin humain, il est très facile de désigner le responsable d'une erreur médicale. Mais ce n'est pas le cas avec des algorithmes qui prennent des décisions. Faut-il tenir l'algorithme responsable ? Le développeur ou plutôt l'entreprise qui a développé cet algorithme et qui tire des profits de son utilisation ? Et si le produit est certifié, n'est-ce pas plutôt l'organisme de certification qui doit être blâmé et juridiquement sanctionné ?

Le questionnement en éthique publique introduit, on le voit bien, une réflexion sur les institutions qui permettent d'offrir des réponses crédibles à un problème moral. Il porte aussi sur le type de société que nous voulons et sur les principes de son organisation. En poursuivant la réflexion sur les *chatbot*s médicaux, on ne peut éluder la question de l'utilité de développer de telles machines intelligentes, de leur intérêt social et humain. On doit en effet se demander s'il est acceptable que des applications intelligentes remplacent des médecins humains même en faisant l'hypothèse qu'elles sont capables de faire des diagnostics précis, voire plus précis que les humains. Que signifie une relation patient-médecin quand le médecin est un *chatbot* ? Que gagne-t-on et que perd-on d'essentiel ? Ce n'est pas une question de type « utilitariste » mais une question qui porte sur la signification de nos relations sociales, sur la reconnaissance de notre vulnérabilité comme patient, sur l'identité humaine. Allons plus loin : investir dans le développement de ce genre d'IA repose sur un choix social éminemment discutable et impliquant donc une discussion collective sur la société que l'on souhaite construire. On peut en effet considérer qu'on devrait améliorer l'accès à un système de santé publique performant et donc investir davantage dans la formation des médecins et dans une organisation équitable de la santé.

[20] Le service de santé publique britannique, le NHS (National Health Service) a ainsi récemment créé une bibliothèque d'applications dans lesquelles on peut avoir confiance (NHS Apps Library). Les applications qui n'offrent pas de garanties suffisantes peuvent être supprimées de la bibliothèque entraînant de sévères conséquences commerciales pour l'entreprise qui vend l'application.

3.4

L'ÉTHIQUE DE L'IA ET LA DÉCLARATION DE MONTRÉAL

Le développement de l'IA et de ses applications met donc en jeu des valeurs morales fondamentales qui peuvent entrer en conflit et provoquer des controverses éthiques, sociales et politiques graves : faut-il développer des applications comme Melody pour diagnostiquer plus rapidement les personnes isolées ou investir autrement dans le système de santé pour que tout le monde puisse consulter un médecin humain ? Il n'y a aucune réponse simple, mais il y a des choix à faire.

La *Déclaration de Montréal* fournit un vocabulaire moral pour repérer les situations sociales problématiques, les analyser et élaborer des réponses pratiques. L'analyse du cas du *chatbot* Melody illustre cette fonction de la Déclaration. Pour saisir l'enjeu du jugement éclairé du patient face au diagnostic, celui de l'attribution de la faute en cas de mauvais diagnostic ou celui de l'accès à un service de santé, la *Déclaration de Montréal* offre un répertoire de valeurs auquel on peut se référer immédiatement : l'autonomie, la responsabilité, l'équité ou la justice. La valeur de vie privée, par exemple, permet de situer le problème de la confidentialité des données du patient.

Le premier objectif de la Déclaration consiste à identifier les principes et les valeurs éthiques qui promeuvent les intérêts fondamentaux des personnes et des groupes. Ces principes appliqués au domaine du numérique et de l'intelligence artificielle restent généraux et abstraits. Pour les lire adéquatement, il convient de garder à l'esprit les points suivants :

1. Bien qu'ils soient présentés sous forme de liste, ils ne sont pas hiérarchisés. Le dernier principe n'est pas moins important que le premier. Mais il est possible selon les circonstances d'attribuer plus de poids à un principe qu'à un autre, ou de considérer qu'un principe est plus pertinent qu'un autre.
2. Bien qu'ils soient divers, ils doivent faire l'objet d'une interprétation cohérente afin d'éviter tout conflit qui empêche leur application. D'une manière générale, les limites de l'application d'un principe sont tracées par le domaine d'application d'un autre principe.
3. Bien qu'ils reflètent la culture morale et politique de la société dans laquelle ils ont été élaborés, ils constituent une base pour une dialogue interculturel et international.
4. Bien qu'ils puissent être interprétés de diverses manières, ils ne peuvent pas être interprétés de n'importe quelle manière. Il est impératif que l'interprétation soit cohérente.
5. Bien que ce soient des principes éthiques, ils peuvent être traduits en langage politique et interprétés de manière juridique.

La Déclaration de principes est suivie d'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 peut couvrir tous les secteurs d'application de l'IA ; ce serait une ambition vouée à l'échec. 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.

La *Déclaration de Montréal* est adressée à toute personne, toute organisation de la société civile et toute compagnie désireuses de participer au développement de l'intelligence artificielle de manière responsable, que ce soit pour y contribuer scientifiquement et technologiquement, pour développer des projets sociaux, pour élaborer des règles (règlements, codes) qui s'y applique, pour pouvoir en contester les orientations mauvaises ou imprudentes, ou encore pour être en mesure de lancer des alertes à l'opinion public quand cela est nécessaire.

Elle s’adresse également aux responsables politiques, élus ou nommés, dont les citoyens attendent qu’ils prennent la mesure des changements sociaux en gestation, qu’ils mettent en place rapidement les cadres permettant la transition numérique pour le bien de tous, et qu’ils anticipent les risques sérieux que présente le développement de l’IA.

Les recommandations qui suivent la Déclaration sont adressées 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.

4. LA DÉMARCHE DE COCONSTRUCTION

4.1 LES PRINCIPES DE LA DÉMARCHE DE COCONSTRUCTION

Pour répondre aux nombreuses interrogations que pose l'usage des machines intelligentes et faire en sorte que l'IA se développe « en bonne intelligence » avec la démocratie, il est nécessaire de recourir à un « surplus » de démocratie et de faire participer le plus grand nombre de citoyens au processus de réflexion sur les enjeux sociaux de l'IA. L'objectif de la démarche de coconstruction est d'ouvrir une discussion démocratique sur la manière dont on doit organiser la société pour faire un usage responsable de l'IA.

Il ne s'agit pas seulement de savoir ce que les individus pensent de telle innovation et de sonder leurs préférences « intuitives »; la coconstruction n'est pas un sondage d'opinion sur des questions du type : « Avez-vous peur que l'IA remplace les juges ? », « Préférez-vous un humain pour vous opérer plutôt qu'un robot ? ». Ce genre de question n'est pas dénué d'intérêt et la méthode du sondage donne des informations importantes aux décideurs publics et offre un précieux matériau de travail aux sciences sociales. Toutefois, si la coconstruction invite à réfléchir collectivement aux enjeux démocratiques, elle demande également l'élaboration de réponses argumentées aux questions pressantes et la formulation de recommandations politiques et juridiques. Le processus de coconstruction leur confère aussi une certaine légitimité démocratique qui crée les conditions d'un débat politique et d'une responsabilisation des décideurs publics, des professionnels et de l'industrie.

C'est tout le sens de la démarche initiée par la Déclaration de Montréal : rendre à la démocratie la compétence de trancher les questions morales et politiques qui concernent la société dans son ensemble. L'avenir de l'IA n'est pas seulement écrit dans des algorithmes, il réside d'abord dans l'intelligence humaine collective.

4.1.1 Les principes d'une bonne participation citoyenne

Dès lors que l'on fait intervenir le public dans un processus de consultation et de participation sur des problématiques sociales controversées, il faut s'assurer que ce processus soit conduit de façon à éviter les risques habituellement associés à l'exercice démocratique. On fait traditionnellement deux objections pour disqualifier le recours au public [21]:

1. L'ignorance : selon cette objection qui est la plus fréquente, le public serait ignorant et n'aurait pas la capacité à comprendre des enjeux techniques qui requièrent un savoir scientifique, une maîtrise des formes logiques de l'argumentation et une connaissance des processus politiques et juridiques.
2. Le populisme : selon cette objection, la participation du public non qualifié peut être l'occasion d'une manipulation démagogique qui flatte les préjugés populaires et peut conduire à l'adoption de propositions déraisonnables, hostiles au progrès social, voire tyranniques à l'égard des minorités.

Si les préjugés et une tendance à l'irrationalité ne peuvent pas être complètement éliminés chez les individus (y compris chez les experts), il est possible de surmonter ces biais de manière collective. Dans des conditions favorables, les individus non experts peuvent participer à des débats complexes sur les problèmes sociaux, comme ceux que présente aujourd'hui la recherche en IA et ses applications industrielles. Nous pouvons identifier 4 conditions nécessaires pour que le processus de coconstruction ne soit pas détourné par les biais cognitifs des participants : la diversité épistémique, l'accès à une information pertinente, la modération, l'itération.

[21] La littérature mettant en cause les compétences politiques de citoyens a connu un regain d'intérêt ces dernières années. Voir entre autres, Jason Brennan, *Against Democracy*, Princeton, PUP, 2016; Ilya Somin, *Democracy and Political Ignorance*, Stanford, SUP, 2013.

A. LA DIVERSITÉ ÉPISTÉMIQUE

Il faut tout d'abord s'assurer que les groupes délibérants manifestent la plus grande diversité interne, en termes de milieu social, de genre, de génération, ou d'origine ethnique. Cette diversité n'est pas seulement requise par l'idée que l'on se fait d'une démocratie inclusive, elle est aussi nécessaire pour augmenter la qualité épistémique des débats. Cela signifie simplement que chacun apporte une perspective différente sur le sujet débattu et enrichit la discussion[22].

B. L'ACCÈS À UNE INFORMATION PERTINENTE

Nous savons cependant que la diversité épistémique ne suffit pas et que si les participants n'ont aucune compétence ou aucune connaissance relative au domaine envisagé, ils ne peuvent produire aucune connaissance nouvelle ni s'orienter dans la discussion. Ils risquent alors collectivement d'amplifier les erreurs individuelles. Il faut donc préparer les participants en leur fournissant une information pertinente et de qualité, à la fois accessible et fiable. La délibération doit donc être précédée par une phase d'information.

C. LA MODÉRATION

Outre le fait de disposer d'une information de qualité, il est nécessaire que les participants raisonnent librement, c'est-à-dire d'abord sans être entravés par des biais cognitifs. On appelle biais cognitifs, les distorsions de la pensée rationnelle par des mécanismes intuitifs. L'un des plus communs et des plus problématiques dans une délibération est le biais de confirmation : on a tendance à n'admettre que les opinions qui confirment nos propres croyances, et à rejeter celles qui vont à l'encontre de ce que nous croyons déjà. Il y a des dizaines de biais cognitifs qui peuvent déformer le cours logique de notre réflexion.

Mais il existe aussi des biais propres à la délibération elle-même, comme la tendance à adopter des positions de plus en plus radicales : si le groupe qui délibère est initialement méfiant à l'égard des innovations en IA, il est probable qu'il y soit tout à fait hostile à la fin du processus de délibération. C'est pour éviter ce genre de résultat mécanique qu'il est important de s'assurer de la diversité épistémique du groupe délibérant et de mettre en place une instance de modération.

Celle-ci ne prend pas nécessairement la forme d'une intervention personnelle par un modérateur. Si nous ne renonçons pas à la modération personnelle, nous croyons pouvoir surmonter les biais de la délibération par d'autres moyens, comme en introduisant des événements imprévus dans les scénarios amorçant les discussions.

D. L'ITÉRATION

Idéalement, nous devrions pouvoir convoquer l'ensemble de la population pour participer à la réflexion sur le développement responsable de l'IA. Cependant, les conditions que nous avons décrites ne peuvent pas être mises en œuvre pour de très grands groupes, encore moins pour une société de plusieurs millions de personnes. Il est donc important de mener la participation citoyenne dans des groupes restreints et de multiplier les rencontres. C'est la phase d'itération de la coconstruction.

[22] Estlund, David M. (2008). *Democratic Authority*: A Philosophical Framework. Princeton University Press.

Les raisons pour procéder ainsi sont techniques, mais peuvent facilement être comprises. Un mathématicien et acteur de la Révolution française, le marquis de Condorcet, avait montré que le jugement des groupes est toujours plus exact que celui des individus pris séparément, et que son exactitude augmente à mesure que le groupe est grand. Mais il y a deux conditions pour que ce soit le cas : il faut que les individus dans le groupe aient plus d'une chance sur deux (1/2) d'avoir raison, et il ne faut pas qu'ils communiquent entre eux (Condorcet craignait à juste titre les risques de manipulation).

Or, dans les très grands groupes, on ne peut pas s'assurer que tous les individus aient la compétence requise et que chacun ait plus d'une chance sur deux d'avoir une opinion adéquate. Permettre la délibération (la communication entre eux) est un des moyens d'augmenter la compétence des participants pourvu qu'elle soit encadrée, comme nous l'avons fait. Certes, cela ne satisfait pas la 2e condition de Condorcet, mais cela permet de garantir la 1re condition. Et pour accroître la qualité des opinions, il convient alors de multiplier les groupes qui délibèrent : puisque l'on ne peut pas accroître la taille du groupe, il faut accroître le nombre de participants en procédant à une itération des sessions de participation[23].

Pour toutes ces raisons, nous avons privilégié la forme de l'atelier de coconstruction qui réunit des citoyens non experts, des experts, des parties prenantes (associations, syndicats, représentants professionnels, entreprises) et des acteurs de la vie politique. Ces ateliers sont organisés selon des formats différents qui sont adaptés aux lieux de délibération et aux disponibilités des participants, et permettent de satisfaire aux conditions d'une participation citoyenne féconde et robuste. Mais il faut noter que le processus d'élaboration de la Déclaration est complexe et repose sur d'autres types de consultations : questionnaires en ligne, rapports et tables rondes d'experts. La Déclaration n'est pas l'enregistrement pur et simple de la parole recueillie dans les ateliers de coconstruction ; elle est le fruit d'une délibération multiple et d'une réflexion qui s'appuie sur les ateliers de coconstruction.

4.1.2 Experts et citoyens

« Pourquoi donner la parole aux citoyens sur des questions éthiques et politiques complexes qui demandent une bonne connaissance des technologies discutées ? Pourquoi ne pas consulter plutôt les experts seulement? » Il y a de nombreuses raisons, mais la plus simple est que l'IA affecte la vie de tous, qu'elle est l'affaire de tous, et que tout le monde doit avoir son mot à dire sur les orientations socialement désirables de son développement.

Même lorsque nous ne sommes pas en présence d'un dilemme moral au sens strict, les questions d'éthique publique ne peuvent être tranchées sans faire des choix qui valorisent certains intérêts moraux au détriment d'autres, sans pour autant les négliger. C'est le résultat du pluralisme des valeurs qui définit le contexte moral et politique des sociétés démocratiques modernes. Il est ainsi possible que l'on valorise le bien-être en contestant la priorité du consentement : pensons à une application médicale qui aurait un accès à des données personnelles pour lequel on n'a pas consenti mais qui permettrait de guérir plus efficacement des maladies graves grâce à ces données. Ce genre de choix éthique et social revient à l'ensemble des membres de notre société démocratique et non à une partie, à une minorité, fût-elle experte.

Le rôle des experts n'est pas de résoudre, à la place des citoyens, les dilemmes éthiques que pose l'IA ni de se transformer en législateurs. À quoi servent les experts alors ? Les experts qui participent au processus de coconstruction de la Déclaration de Montréal n'ont pas l'intention de raisonner à la place des citoyens pour proposer un cadre éthique et légal que ces derniers se contenteraient de valider. Pour penser les enjeux éthiques et sociaux complexes de l'IA, l'expertise doit être au service de la réflexion citoyenne.

[23] Estlund (2008); Landemore, Hélène. (2013). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton University Press.

Parfois les éthiciens donnent l'impression de vouloir donner des leçons de morale, de connaître les réponses aux questions épineuses que se pose le public, et même de pouvoir régler d'avance les défis de demain. Il est important de préciser leur rôle. Dans le processus de coconstruction, les éthiciens ont trois tâches à la fois modestes et cruciales :

> S'assurer des conditions favorables de la participation citoyenne

> Clarifier les enjeux éthiques qui sous-tendent les controverses autour de l'intelligence artificielle

> Rationaliser les arguments défendus par les participants en leur indiquant les arguments que l'on sait erronés ou biaisés et en leur expliquant les raisons pour lesquelles ils sont erronés.

Le rôle des éthiciens est donc celui d'un accompagnement éclairé[24]. Les experts dans les autres domaines de recherche (en informatique, en santé, en sécurité, en droit, etc.) ont également un rôle d'accompagnement en fournissant aux participants les informations les plus utiles et les plus fiables sur l'objet de la controverse : Comment fonctionne un algorithme qui apprend à établir des diagnostics ? Le médecin peut-il être remplacé par un robot programmé pour le diagnostic ? Quelles sont les protections que nous pouvons opposer aux tentatives de piratage de nos données médicales ? Etc.

Toutefois, il faut bien admettre que les experts eux-mêmes manifestent parfois des biais cognitifs importants. Ils peuvent se montrer trop optimistes ou trop pessimistes sur les nouvelles technologies qu'ils connaissent bien ; ils ont également tendance à être trop confiants dans leur jugement, en particulier quand ils estiment être en mesure de prédire les évolutions de leur domaine de recherche, les changements sociaux, etc. C'est en les faisant participer comme citoyens aux ateliers de coconstruction qu'on réduit les biais propres à l'expertise ainsi que les effets d'autorité que produit l'asymétrie de savoir avec les autres participants.

Les ateliers de coconstruction sont des lieux de participation qui permettent de donner des orientations au développement socialement souhaitable de l'IA, d'innover par des propositions qui font bouger les cadres d'analyse admis. Cet apport essentiel de la délibération citoyenne est ensuite analysé et approfondi par des comités de travail constitué d'experts de différents milieux (chercheurs, professionnels). Ce travail d'approfondissement et de rédaction de recommandations suit les orientations définies par la délibération et reste fidèle aux propositions issues des ateliers de la coconstruction.

4.2

LA MÉTHODOLOGIE DES ATELIERS DE COCONSTRUCTION

La version préliminaire de la *Déclaration de Montréal sur l'IA responsable*, présentée le 3 novembre 2017 lors du Forum IA responsable, sert d'assises au processus de coconstruction. Schématiquement, après avoir statué sur le « quoi ? » (quels principes éthiques souhaitables devraient être rassemblés dans une déclaration sur l'éthique de l'intelligence artificielle ?), il s'agit dans cette nouvelle phase d'anticiper avec les citoyens et parties prenantes comment des controverses éthiques pourraient surgir dans les prochaines années à propos de l'IA (dans les secteurs de la santé, de la justice, de la ville intelligente, de l'éducation et de la culture, du monde du travail et des services publics), pour imaginer ensuite comment on pourrait y répondre (par exemple, par un dispositif comme une certification sectorielle, un nouvel acteur-médiateur, un formulaire ou une norme, par une politique publique ou un programme de recherche).

L'objectif de la démarche de coconstruction et de ses ateliers est de mettre à l'épreuve les principes de la Déclaration de Montréal à l'aide de scénarios prospectifs. Ultimement, le processus permettra de préciser les enjeux éthiques sectoriels, et de formuler des recommandations prioritaires auprès de la communauté IA.

[24] Weinstock, Daniel M., *Profession éthicien*, Montréal, Presses de l'Université de Montréal, 2006.

DÉCLARATION DE MONTRÉAL IA RESPONSABLE

(3 novembre 2017)

7 principes pour un déploiement responsable de l'IA dans la société

>

Ateliers de coconstruction

(février-octobre 2018)

Scénarios sectoriels 2025 et principes de la Déclaration de Montréal IA responsable

v

Enjeux éthiques sectoriels

v

Recommandations prioritaires

>

Recommandations aux gouvernements, aux parties prenantes, aux chercheurs

DÉCLARATION DE MONTRÉAL IA RESPONSABLE
(décembre 2018)

10 principes pour un déploiement responsable de l'IA dans la société

Plus de 10 ateliers de coconstruction ont été organisés de février à octobre : des cafés citoyens de 3 heures dans des bibliothèques publiques, et deux grandes journées de coconstruction avec des citoyens, des experts et des parties prenantes variées (à la SAT à Montréal, au Musée de la civilisation à Québec et enfin au Centre culturel canadien[25]).

Le choix d'organiser des cafés citoyens dans les bibliothèques publiques est explicitement lié à la dynamique de réinvention actuelle de ces services publics au Québec et au Canada[26]. En passant du modèle de l'espace de prêt documentaire à celui de la « bibliothèque lieux » inclusive et cherchant à renforcer les capabilités de tous les citoyens (ex. avec des services de littératie numérique, de soutien aux citoyens, des espaces de discussion et de médiation culturelle, le prêt d'outils et la création de fab labs), les bibliothèques publiques auront très certainement un rôle clé à jouer dans le déploiement responsable de l'IA au Québec et au Canada.

Les journées de coconstruction se sont déroulées dans des lieux emblématiques (Société des arts technologiques à Montréal, Musée de la civilisation à Québec) et ont notamment mis l'accent sur la rencontre entre les parties prenantes et les disciplines très variées qui doivent collaborer pour imaginer un déploiement responsable de l'IA dans la société québécoise.

[25] Nous remercions l'Ambassade du Canada à Paris d'avoir rendu possible la tenue de cet atelier à Paris qui a eu lieu le 9 octobre 2018.

[26] Christophe Abrassart, Philippe Gauthier, Sébastien Proulx et Marie D. Martel, *Le design social : une sociologie des associations par le design ? Le cas de deux démarches de codesign dans des projets de rénovation des bibliothèques de la Ville de Montréal*, Lien social et Politiques, 2015, n° 73, p. 117-138

4.3

ORIGINALITÉ DE LA DÉMARCHE DE COCONSTRUCTION

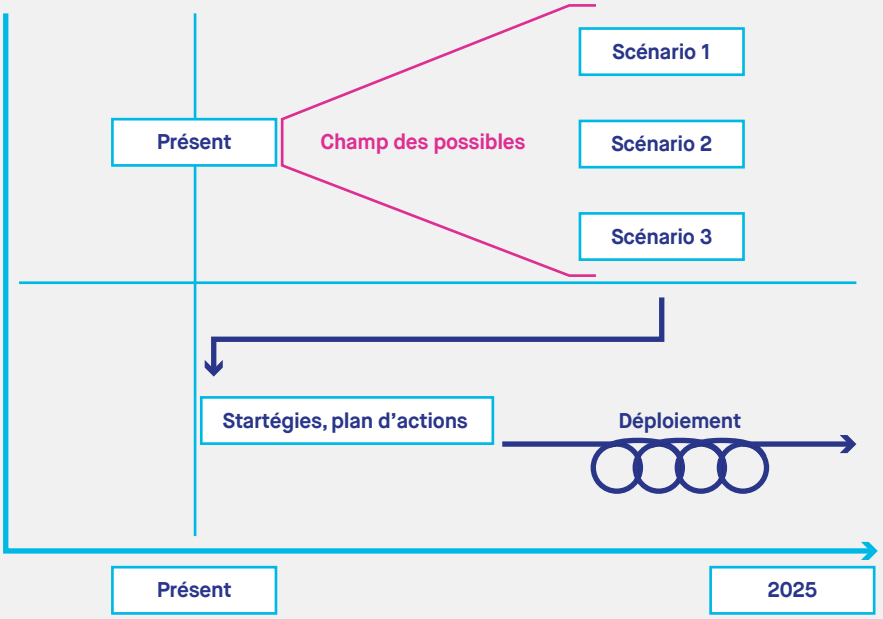
Au regard des autres initiatives en éthique de l'IA actuellement en cours dans le monde, cette démarche de coconstruction présentera en particulier trois dimensions originales et innovantes :

> Tout d'abord, le recours aux méthodes de prospective stratégique (« foresight »), avec des scénarios sectoriels en 2025 exemplifiant par de courts récits comment des controverses éthiques sur l'IA pourraient surgir ou s'amplifier dans les prochaines années (dans les secteurs de la santé, de la justice, de la ville intelligente, de l'éducation et de la culture, du monde du travail). Ces scénarios 2025, qui présentent une variété de situations possibles face à un avenir très ouvert, seront utilisés comme déclencheurs de débats, pour identifier, préciser ou anticiper des enjeux éthiques sectoriels sur le déploiement de l'IA dans les prochaines années. Ces discussions à l'horizon 2025 permettront ensuite de formuler rétrospectivement des recommandations concrètes pour 2018-2020, pour nous diriger vers des situations collectivement souhaitables.

Figure 4 : La prospective stratégique : une démarche en trois temps

LA PROSPECTIVE STRATÉGIQUE :
une démarche en trois temps (Lab Ville Prospective)

1. Exploration collective : Quels enjeux éthiques sectoriels de l'IA en 2025 ?
2. Rétrospection : Quelles recommandations pour 2018-2020 ?
3. Action collective : Expérimentation collective et déploiment ?



> Ensuite, le recours à des méthodes d'animation de design participatif en « forum hybride »[27] pluridisciplinaire, incluant les citoyens et les parties prenantes, dans un contexte d'incertitude partagée face aux futurs possibles (pour approfondir un scénario, concevoir des dispositifs de réponse à un risque éthique, proposer un complément à la Déclaration de Montréal en cas d'enjeu orphelin, i.e. sans principe éthique correspondant).

> Enfin, une attention aux « biais de paradigmes » qui ont des effets de cadrage très puissants dans la manière de poser les problèmes (ex. aborder les enjeux éthiques de la voiture autonome uniquement sous l'angle du dilemme du tramway comme le propose l'équipe de l'expérience *Moral Machine* du MIT) et dans le cadre du paradigme de la « vitesse-distance » en design des transports), dans le but d'assurer un pluralisme des enjeux et de rendre visibles des situations encore inconnues ou très émergentes dans un contexte de changement rapide.

Cette démarche de coconstruction vise d'une manière générale à élaborer une *trajectoire apprenante* pour concevoir, au fil des événements, une trousse d'animation reproductible, conviviale et adaptable, qui pourra être publiée en « open source » à l'issue de la démarche de coconstruction.

Le détail des cafés citoyens et des journées de coconstruction se trouve en annexe du rapport.

4.4

CAFÉS CITOYENS EN MARGE DES BIBLIOTHÈQUES

Il faut également mentionner l'implication de deux étudiants en philosophie de l'Université de Montréal, Pauline Noiseau et Xavier Boileau, qui ont organisé de février à avril 2018, plusieurs cafés citoyens dans des lieux publics autres que les bibliothèques, et dont la formule était davantage axée sur les discussions libres autour d'un enjeu de l'IA. Les modérateurs ont utilisé des scénarios très courts, et animé des séances de 2 heures. Ces séances ont constitué des moments forts de délibération avec des citoyens qui ne demandent qu'à participer davantage aux débats publics, mais qui sont rarement sollicités. Ainsi, un café citoyen à la Maison d'Haïti, le 25 avril 2018, a permis à des jeunes scolarisés au secondaire et à des retraités du quartier Saint-Michel de Montréal-Nord d'échanger autour des enjeux de l'IA. À partir d'un scénario sur l'IA des objets connectés domestiques (un réfrigérateur intelligent), cette séance a notamment suscité des réflexions originales sur la cuisine comme activité humaine relationnelle posant des enjeux d'authenticité, de lien affectif (la « touche d'amour ») et d'habileté sociale, enjeux qui n'étaient pas ressortis des autres types de consultation à partir du même scénario.

4.5

PORTRAIT DES PARTICIPANTS

Le recrutement de citoyens, d'experts et de professionnels de différents secteurs du marché du travail a permis d'avoir une diversité de participants pour la coconstruction. Les facultés universitaires, ainsi que les centres de recherche interuniversitaires et leurs réseaux, ont permis de rejoindre un nombre important d'acteurs impliqués dans le développement et l'utilisation de l'IA.

[27] Callon, Lacoumes, Barthe, Agir dans un monde incertain. Essai sur la démocratie technique, Paris, Le Seuil, 2001

Pour rejoindre le grand public, les réseaux sociaux et sites web des différents partenaires ont joué un rôle important, bien que ce soit les efforts de recrutement locaux de chaque bibliothèque impliquée qui furent les plus déterminants.

Fait à noter, il y a eu une quasi-parité hommes-femmes dans les ateliers. Une grande majorité des participants possèdent une éducation post-secondaire et se trouve dans la tranche d'âge 19-34 ans.

Figure 5 : Proportion hommes-femmes ayant participé aux ateliers de coconstruction

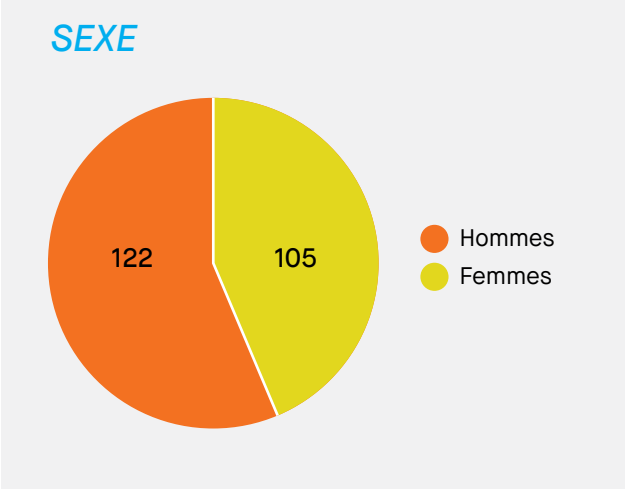


Figure 6 : Les participants aux ateliers de coconstruction par tranches d'âge

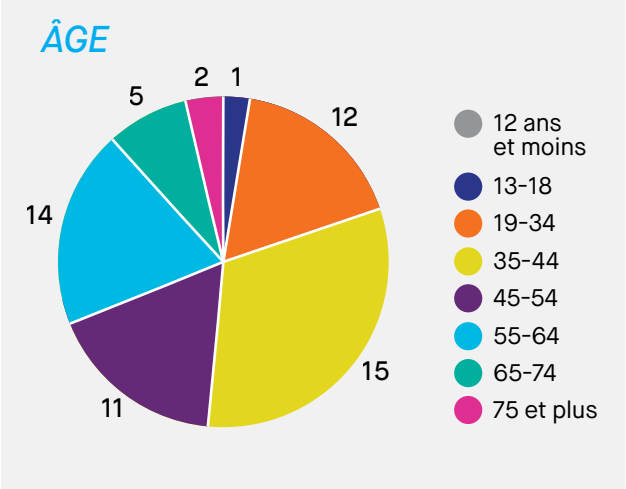


Figure 7 : Répartition des répondants aux cafés citoyens et aux journées de coconstruction par niveau de scolarité atteint

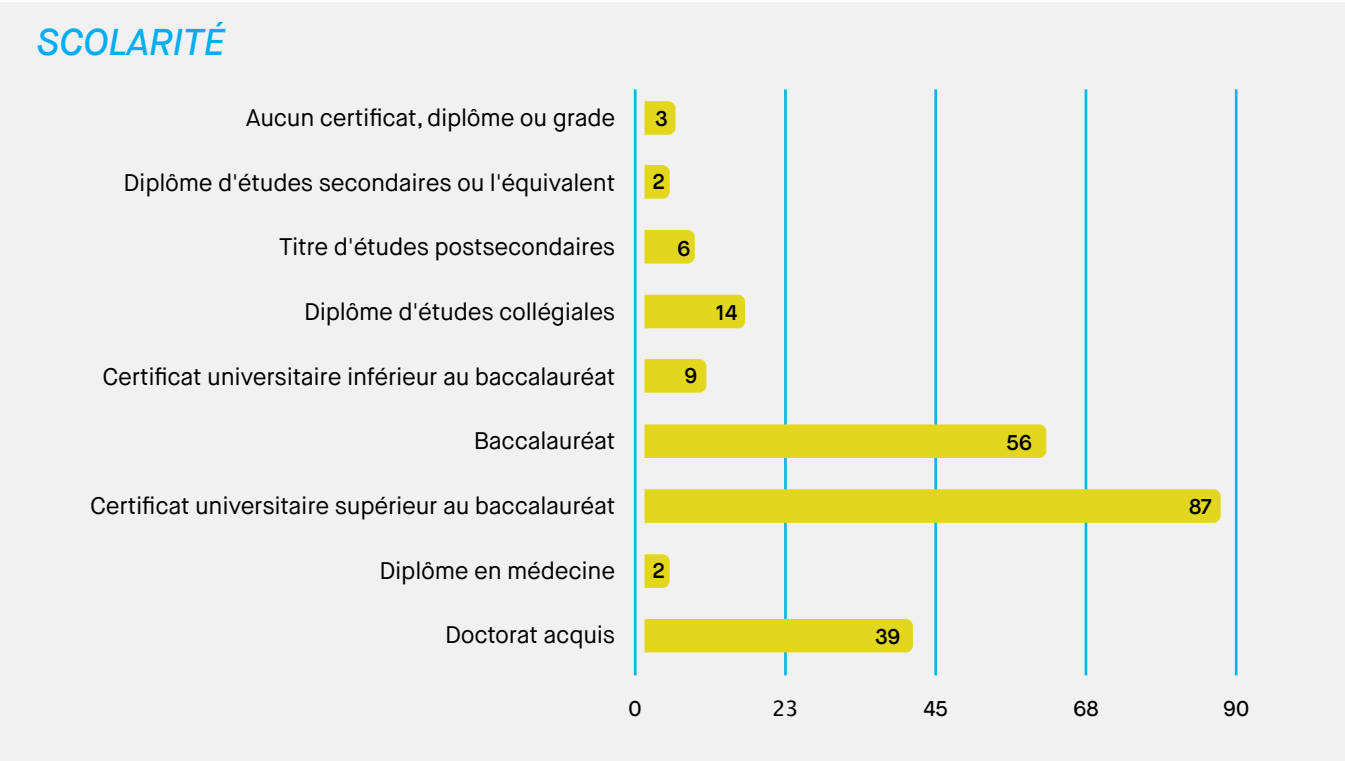
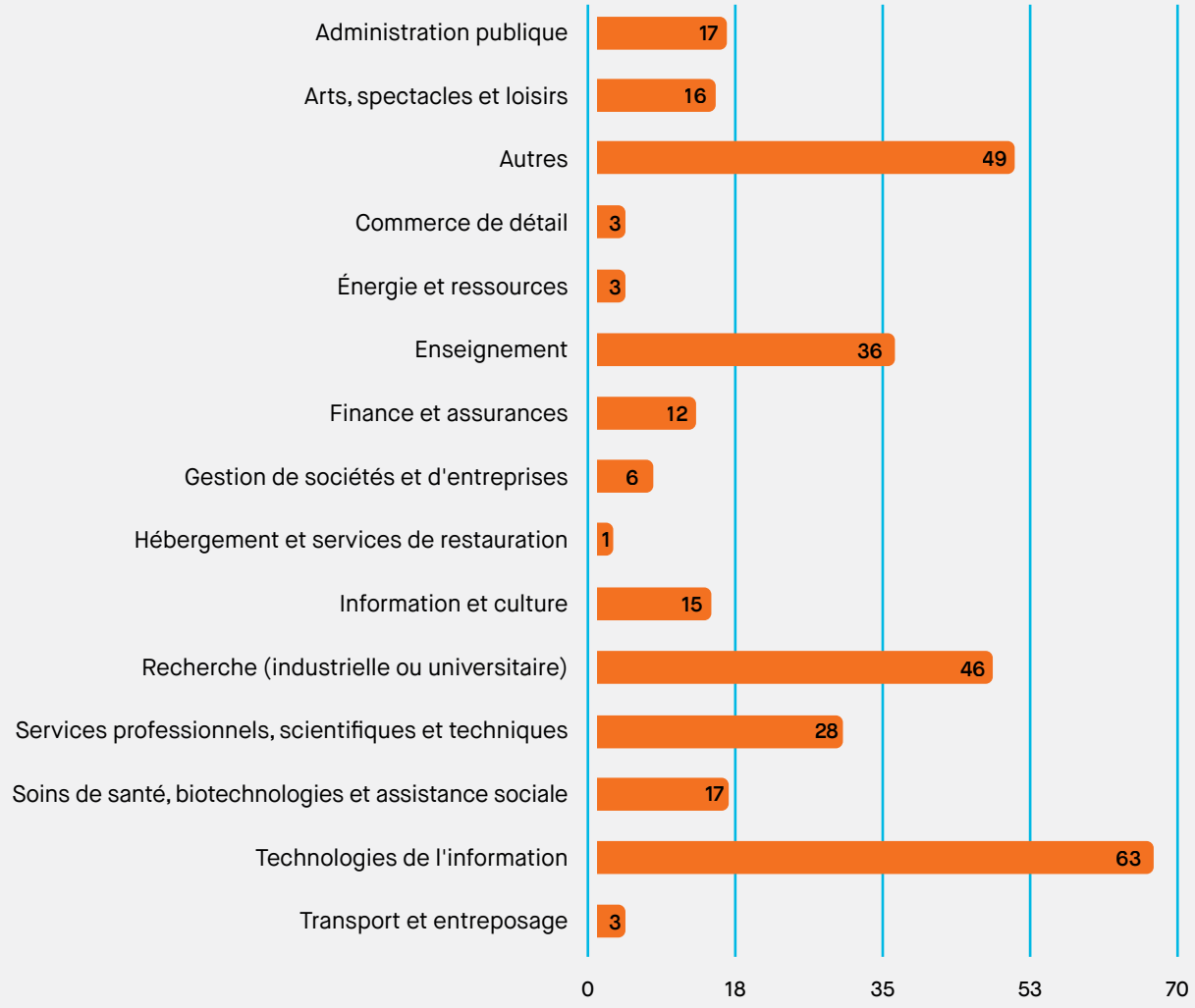


Figure 8 : Répartition des répondants aux cafés citoyens et aux journées de coconstruction par secteur d'activité

SECTEURS D'ACTIVITÉ

34% des répondants ont indiqué plus d'un secteur d'activité



5. PARCOURS DÉLIBÉRATIFS DANS LES ATELIERS : exemples de deux secteurs : ville intelligente et monde du travail

5.1

LES PARCOURS DÉLIBÉRATIFS

Comment se sont déroulées les discussions et les délibérations dans les ateliers de coconstruction ? Quelles réflexions ont-elles suscitées ? Quels ont été les grands jalons de la discussion pour parvenir à des propositions d'encadrement de l'IA ? Cette section du document présente en détail certains faits saillants des délibérations entre les participants, où chacun a pris soin de préciser les raisons, les principes et les valeurs justifiant sa position sur le scénario prospectif proposé comme point de départ, que ce soit pour exprimer un accord, un désaccord, une nuance ou un nouveau questionnement. En un mot, pour faire ce que la sociologie pragmatique a qualifié de justification.

Pour illustrer ce travail, le parcours de deux équipes représentant deux secteurs parmi les cinq abordés dans la coconstruction a été choisi :

une table de citoyens ayant traité de la voiture autonome (secteur ville intelligente) et une table de chercheurs et experts ayant traité de l'impact de l'IA sur l'emploi dans les entreprises (secteur monde du travail).

Pour formuler ses propositions, chaque équipe a parcouru trois étapes où se sont succédées la génération d'idées puis la délibération sur ces idées :

Première étape : la formulation d'enjeux éthiques et sociaux sectoriels en 2025 (en croisant les principes généraux de la *Déclaration de Montréal* et des situations d'usagers en 2025 décrites dans les scénarios déclencheurs de débats) : la formulation individuelle d'enjeux (sur des *Post-it*) a ensuite été approfondie lors d'une discussion collective d'où est ressortie une sélection de trois enjeux prioritaires.

Deuxième étape : la formulation de recommandations à mettre en place dès 2018-2020 pour préparer un déploiement responsable de l'IA au Québec : de la formulation de recommandations au choix de quelques mesures pour la « Une » du Journal.

Troisième étape : la mise en récit du lancement d'une première recommandation en 2020 (la « Une » du Journal) pour prendre la mesure du « temps de l'action collective » avec ses contraintes organisationnelles : de la formulation d'idées à leur synthèse ordonnée dans un récit.

Soulignons qu'entre ces étapes et micro-étapes du parcours délibératif, la « nature » des idées générées varie : certaines sont des intuitions individuelles (lorsqu'au début de l'exercice, les participants formulent sur des *Post-it* plusieurs enjeux sectoriels), d'autres résultent d'une discussion collective (où chacun justifie son point de vue) et d'autres enfin, sont le résultat d'une hiérarchisation effectuée par le groupe (lorsqu'en fin d'étape les participants sélectionnent trois enjeux clés à inscrire sur l'affiche de synthèse).

On retrouve ainsi dans le fonctionnement de ces ateliers prospectifs trois propriétés des dispositifs délibératifs soulignés par Blondiaux et Sintomer dans leur article *L'impératif délibératif*[28] : rendre possible l'imagination de solutions nouvelles dans un monde incertain ; permettre une progression en généralité et viser des consensus ou des « désaccords délibératifs » dans une société marquée par le pluralisme des valeurs ; et enfin, donner une source factuelle et normative de la légitimité par l'inclusion de tous à ces délibérations.

[28] Blondiaux L. et Sintomer Y., *L'impératif délibératif*, revue Politix, 2002, p. 25-26

5.1.1 Secteur ville intelligente : la voiture autonome (VA) et le juste partage de la rue

Résumé du scénario 2025 de départ. En 2025, les premières VA circulent à Montréal et une controverse sur le partage de la rue et de l'espace public apparaît. Des voies sont maintenant réservées aux VA et protégées par des barrières, pour qu'ils puissent rouler sans risque d'accident à une allure modérée, mais fluide (50 km/h). Les VA peuvent aussi rouler ailleurs, mais à des vitesses très lentes (25 km/h). Les manifestants pour la mobilité active (marche, vélo) perturbent le fonctionnement de ces voies protégées, sachant que les algorithmes des VA sont réglés en mode « altruiste », pour protéger les personnes extérieures.

L'objectif de ce scénario était d'ouvrir une discussion sur les enjeux éthiques de la VA à partir d'une situation restituant la densité et la complexité de la ville : des vitesses lentes et différentes, la fluidité comme critère prioritaire à la vitesse, des barrières de protection pour la sécurité, la rue comme espace partagé entre des usages concurrents.

Le parcours délibératif présenté est issu d'une table tenue en 3 h dans une bibliothèque publique de Montréal, avec huit citoyens intéressés par les nouvelles technologies et ayant par ailleurs des pratiques de mobilité active en famille (vélo, marche). Partant de ce scénario de 2025, la discussion a débouché sur la formulation d'une initiative présentée en « Une » du Journal de l'IA responsable du 13 mars 2020 : « Premier atelier de littératie en mobilité autonome ». Quel a été le parcours délibératif de ce groupe pour mener à cette proposition originale ? Quels ont été ses moments marquants ? Comment se sont enrichies les idées à chaque étape ? Nous présentons, en les commentant, certains moments significatifs du parcours suivi par cette équipe.

Premier moment délibératif : FORMULATION D'ENJEUX ÉTHIQUES EN 2025

De nombreuses interrogations rédigées sur des *Post-it* ont été soumises par les participants en relation avec différents principes de la *Déclaration de Montréal* :

LE PRINCIPE D'AUTONOMIE

« Est-ce que les humains vont devenir trop dépendants lors de leur déplacement ? », « Est-ce que la liberté de mouvement sera limitée par l 'IA ? », « On délègue beaucoup de micro-décisions à des IA et systèmes interconnectés au détriment de l'humain. »

LE PRINCIPE DE BIEN-ÊTRE

« Beaucoup moins de place pour la spontanéité avec les VA. » , « Quel sera le développement des quartiers par rapport aux axes routiers des VA ? » , « Est-ce que les données des déplacements influencent l'urbanisation des villes ? »

LE PRINCIPE DE DÉMOCRATIE ET DE JUSTICE

« Quelle est la différence d'aménagement des axes de déplacement dans des quartiers populaires par rapport aux quartiers aisés ? », « Est-ce que seuls les mieux situés bénéficient de la fluidité du trafic ? »

LE PRINCIPE DE VIE PRIVÉE

« Pourra-t-on retracer tous les déplacements des gens ? », le principe de responsabilité : « Qui a la responsabilité de l'accident ? », ou de sécurité : « Possibilité de *hacker* des flottes de véhicules ? », ce dernier principe étant proposé par les participants, en complément de ceux de la déclaration.

Plusieurs discussions approfondies ont ensuite eu lieu, les participants rebondissant sur les premières idées pour en générer d'autres sur la spontanéité et la liberté des trajets, sur la sécurité des données personnelles et leur gestion par un organisme centralisé, sur la question du réglage des algorithmes et de la possibilité de les détourner.

Puis, après près de 45 minutes de discussion, les participants ont sélectionné, à l'aide de pastilles colorées, des regroupements d'enjeux éthiques pour 2025 leur semblant prioritaires. Les votes des participants à l'aide de pastilles colorées positionnées sur le tableau des *Post-it* et des idées discutées, ont convergé sur les idées associées à quatre principes de la *Déclaration de Montréal*, dont deux ont été regroupés : sécurité, justice, et bien-être et autonomie.

Tableau 1 : Ville intelligente, Premier moment délibératif : formulation d'enjeux éthiques en 2025

Enjeux éthiques 2025	**1**	**2**	**3**
Description	Facilité de piratage du système centralisé. Dilemme : fluidité collective-vulnérabilité du système	Risque d'exclusion sociale Typologie des réglages par classe sociale (ex : trajet quartier pauvre-réglages VIP)	Perte de spontanéité des trajets, perte d'autonomie et de liberté de mouvement, et géolocalisation.
Principes associés	Sécurité	Justice	Bien-être et autonomie

Cette sélection d'enjeux prioritaires par l'équipe est originale : si l'enjeu de sécurité, et ceux de responsabilité et de vie privée sont souvent invoqués dans les études et les débats sur les VA, ceux de justice, de bien-être et d'autonomie le sont beaucoup moins.

Deuxième moment délibératif : PROPOSITIONS D'ENCADREMENT DE L'IA POUR 2018-2020

Pour répondre à ces enjeux, l'équipe a poursuivi ses discussions en essayant de réfléchir ensemble aux quatre principes associés. Plusieurs propositions d'encadrement de l'IA ont été formulées par les participants. Nous en présentons ici trois (sur six), qui permettent de suivre le cheminement des idées jusqu'à la formulation de la « Une » du Journal.

Tableau 2 : Ville intelligente, Deuxième moment délibératif : propositions d'encadrement de l'IA pour 2018-2020

Propositions d'encadrement en 2018-2020	**1**	**2**	**3**
Description	Formation à la vigilance collective (ex. du permis de conduire)	Comité multipartite qui gère démocratiquement les incidents, injustices et autres enjeux ; le comité doit être décisionnel	Évaluation du plan d'urbanisme pendant la période de transition
Catégories d'instrument	Nouvelle formation	Nouvel acteur institutionnel	Processus de planification participative

Ces propositions, qui dénotent une véritable créativité institutionnelle (au-delà des exemples d'instruments très généraux donnés dans le livret du participant) se situent dans la lignée des enjeux identifiés à l'étape précédente, mais présentent aussi un enrichissement des idées (ce ne sont pas de simples déductions d'instruments adaptés à partir d'un cas éthique identifié). L'idée de formation à la vigilance et celle de la participation à la décision collective (par un comité multipartite et une planification ouverte) conduisent en effet à des propositions de renforcement des capacités et des formes de démocratie locale.

Troisième moment délibératif : ÉCRITURE DE LA « UNE » DU JOURNAL EN 2020

Ces mesures ont ensuite été mises en récit de la manière suivante dans l'affiche. La « Une » du Journal de l'IA responsable du 13 mars 2020 formulée par l'équipe était la suivante :

« PREMIER ATELIER DE LITTÉRATIE EN MOBILITÉ AUTONOME »

« Le réseau des bibliothèques publiques du Québec a mis en place un programme de formation sur l'utilisation des véhicules autonomes. Au programme : la vigilance collective ; le code de déontologie ; comment s'impliquer dans le comité décisionnel de la ville ; l'usage partagé de la rue entre piétons, vélos, VA, camions ; l'explication des règlements ; des séances d'essais ; la question du réglage des algorithmes. »

Cette « Une » du Journal, qui a été formulée à l'issue d'une discussion entre les participants, contribue à nouveau à la progression des idées. En effet, le principe d'un atelier de **« littératie de la mobilité autonome » permet une création de sens inédite** en intégrant les différentes recommandations formulées à l'étape précédente et en élargissant le point de vue pour parler de mobilité autonome et non simplement de VA (donc en permettant la possibilité de trajets multimodaux autonomes). Cette « Une » présente également un **dispositif d'action collective** avec une cible de progrès (la formation et les capacités des citoyens, la possibilité de participer au comité décisionnel des villes sur le déploiement des VA) et une organisation (un déploiement dans les bibliothèques publiques du Québec, qui sont en pleine transformation actuellement pour devenir des tiers-lieux de services culturels pour tous les citoyens).

Le résultat de cette table est particulièrement intéressant car il permet d'envisager la question éthique des véhicules autonome sous l'angle de l'autonomie et de la justice sociale dans la ville, et non exclusivement selon la problématique de la responsabilité dans un scénario d'accident, comme le propose par exemple l'initiative *Moral Machine* du MIT à partir du dilemme moral du tramway[29].

[29] Site du MIT : http://moralmachine.mit.edu

5.1.2 Secteur du monde du travail : une restructuration socialement responsable ?

Résumé du scénario 2025 de départ.
En 2025, de nombreuses entreprises utilisent l'IA dans leurs outils de gestion. C'est le cas d'une entreprise de logistique écologique qui doit investir massivement dans l'IA et la robotisation pour maintenir sa compétitivité. Tri des colis, routage, suivi administratif, calcul des bilans de carbone des trajets, camions autonomes électriques : au total un tiers des postes de l'entreprise pourraient être supprimés. L'entreprise, qui est très engagée socialement, voudrait faire cette restructuration de manière socialement responsable, par exemple en créant une Coop de traitement des *data* pour réembaucher un maximum de salariés, indépendamment des grandes entreprises en place. Y parviendra-t-elle à temps ?

L'objectif de ce scénario était d'ouvrir une discussion sur les enjeux éthiques et sociaux du processus de changement provoqué par l'IA que vont rencontrer les milliers de PME québécoises, ainsi que les grandes entreprises, dans la décennie 2020-2030.

Le parcours délibératif présenté dans cette section est issu d'une table tenue lors d'une journée complète à Montréal regroupant près de dix chercheurs et experts variés travaillant sur les mutations du monde du travail, la participation sociale et la responsabilité sociale des entreprises ou encore pour un syndicat. Une citoyenne ayant préalablement participé à un atelier dans une bibliothèque publique était également présente à cette table.

Partant de ce scénario en 2025, le travail de cette équipe a débouché sur la formulation d'une initiative présentée en « Une » du Journal de l'IA responsable du 18 février 2020 : **« Premières mesures du comité interministériel mixte sur la transition numérique responsable »**. Comme dans le cas précédent sur la voiture autonome, quel a été le parcours délibératif de ce groupe pour aboutir à cette proposition ? Quels ont été ses moments marquants ? Comment se sont enrichies les idées à chaque étape ? Nous présentons, en les commentant, certains moments significatifs du parcours suivi par cette équipe.

Premier moment délibératif : FORMULATION D'ENJEUX ÉTHIQUES EN 2025

De nombreux *Post-it* ont été rédigés dans la première partie de l'atelier en matinée par les participants. En voici un extrait et un aperçu à travers quelques formules issues des *Post-it* et du tableau de leur regroupement par principes de la *Déclaration de Montréal*.

Certains enjeux formulés ont été associés à différents principes de la *Déclaration de Montréal* :

LE PRINCIPE DE BIEN-ÊTRE

« Que doit-on favoriser ? L'entreprise ou la société ? », « Adopter différentes perspectives sur le bien-être : individuel (le salarié), le développement collectif et social, le développement économique (PME) », « Quels sont les idéaux de performance quand le robot ou le *cobot* ne se fatigue pas à la différence de l'humain ? », « Quels sont les aspects positifs possibles : renforcement professionnel, par ex. en médecine, une baisse de la pénibilité sur certains postes », « Quels sont les nouvelles formes de travail et de protection avec le travail-loisir ? »

LE PRINCIPE D'AUTONOMIE

« Quels parcours professionnels et de vie ? Peut-on ne pas réorienter sa carrière en fonction de l'IA ? Avec quelles conséquences ? », « autonomie collective : pour l'anticipation collective et critique du discours de l'urgence de l'adaptation »

LE PRINCIPE DE RESPONSABILITÉ

« Qui est tenu responsable de ces changements ? », « La responsabilité éthique et sociale de la transition est-elle individuelle – chaque entreprise – ou collective – la société, le gouvernement ? », « Quel financement pour la transition ? » ; « Comment aligner l'impératif de rentabilité et la responsabilité dans un contexte d'urgence ? »

LE PRINCIPE DE CONNAISSANCE

« Quelle collaboration entre humains et robots ? Charge de travail, santé-sécurité, formation, acceptabilité, cybersécurité », « Comment sont collectées les données dans un contexte où ce travail est principalement opéré par les entreprises privées (GAFAM) ? », « Comment ne pas figer les gens dans des classes ? », « Quelles sont les possibilités de données partagées ? », « Quel est l'impact sur le système éducatif ? »

LE PRINCIPE DE JUSTICE

« Quelle indépendance face à la concentration de pouvoir des GAFAM ? », « Quelle redistribution sociale des bénéfices de l'IA ? », « Est-ce que les gains de productivité par l'IA et l'industrie 4.0 permettront de financer la transition sociale si les entreprises pratiquent l'évitement fiscal ? », « Quelle équité en cas de partage et codification des connaissances tacites des salariés pour les transformer en *data* ou nourrir la robotisation ? », « A-t-on le choix, en tant que travailleur, de ne pas divulguer ces informations ? », « Sur quels critères va-t-on choisir ceux qui seront remplacés et ceux qui seront formés ? », « Quel accès à la protection sociale de demain ? », « Quels accès aux droits, comme celui d'association, avec les nouvelles organisations du travail ? »

LE PRINCIPE DE DÉMOCRATIE

« La précarisation est-elle une fatalité alors qu'on peut anticiper la transition ? », « la vision à court terme politisée par opposition à la vision à long terme », « l'obscurcissement des processus décisionnels », « des risques de biais dans les ensembles d'apprentissage des algorithmes », « la nécessité d'un débat démocratique ».

Nous pouvons souligner ici que la typologie des principes de la *Déclaration de Montréal* sur l'IA responsable a bien fonctionné pour donner des balises à la discussion, et que les participants ont même proposé des problématiques originales sur certains principes : la nécessité d'aborder le **bien-être** et la r**esponsabilité de la transition** de différents points de vue (individuel et collectif) ; le rapport au temps social, avec l'opposition entre l'anticipation collective et un discours opaque de l'urgence, comme condition de notre **autonomie collective** et de notre exercice de la **démocratie** (le manque de temps empêchant le travail démocratique informé); une forte exigence de **justice** sur la redistribution sociale des bénéfices de l'IA, notamment au sujet de l'équité accompagnant la codification, et donc l'automatisation possible, des compétences des salariés.

Après une bonne heure de discussion, les participants ont sélectionné, à l'aide de pastilles colorées des regroupements d'enjeux éthiques en 2025 leur semblant prioritaires. Les votes étant répartis assez également sur les différents enjeux, jugés aussi importants par le groupe, la formulation des trois enjeux prioritaires pour l'affiche a aussi relevé d'un exercice de synthèse des idées discutées dans cette première partie de l'atelier (voir tableau ci-dessous).

Tableau 3 : Monde du travail, Premier moment délibératif : formulation d'enjeux éthiques en 2025

Enjeux éthiques 2025	**1**	**2**	**3**
Description	Trop de concentration de pouvoir (cf. GAFAM) ne permettant pas : - Le partage équitable des bénéfices de l'IA - L'entrée de nouveaux joueurs (nouveaux modèles d'affaire de type COOP) - La diminution des inégalités (littératie)	Déterminisme technologique, fatalité (« Société boîte noire ») et urgence : au lieu de prendre le temps de faire un débat démocratique informé et participatif sur les nouveaux risques sociaux, les modèles de développement social, les idéaux de performance, l'expérience de travail.	Définir le bien commun et le type de responsabilité collective dans la transition numérique Par exemple : quelles parties prenantes ? L'entreprise seule ? L'État? Les syndicats ? Le système éducatif ?
Principes associés	Justice et indépendance	Démocratie, connaissance, et autonomie collective	Bien-être et responsabilité

Deuxième moment délibératif : PROPOSITIONS D'ENCADREMENT DE L'IA POUR 2018-2020

Pour répondre à ces enjeux, l'équipe a poursuivi ses discussions en après-midi avec un nouveau tour de table menant à la rédaction par les participants de propositions d'encadrement de l'IA sur des *Post-it*, ce qui a conduit à de nombreuses propositions qui ont été discutées une par une collectivement. Le tableau ci-dessous en présente un extrait (six propositions sur plus de dix formulées par le groupe), afin de suivre le cheminement des idées jusqu'à la formulation de la « Une » du Journal.

Tableau 4 : Monde du travail, Deuxième moment délibératif : proposition d'encadrement de l'IA pour 2018-2020

Propositions d'encadrement en 2018-2020	**1**	**2**	**3**	**4**	**5**	**6**
Description	**Renforcer la littératie numérique pour tous.** Avec un référenciel de compétences, les bibliothèques publiques, les écoles, en milieu de travail. En traitant la question de l'analpha-bétisme et du «non-recours» des citoyens.	**Comité permanent interministériel mixte sur l'IA, exécutif à côté du PM.** À l'interface des thèmes de l'économie, de l'emploi, de l'éducation et de la culture (cf. Stratégie numérique)	**Fonds d'assurance numérique sur l'IA pour permettre de se former et s'adapter.** Exemple de dispositif : le Régime d'assurance parentale 50 semaines, qui peut aussi inspirer un revenu minimum contre la précarisation	**Incitatifs sur les nouveaux modèles d'entreprises de traitement des *datas*** Exemple : COOP pour rompre l'isolement de travailleurs autonomes opérant sur les *data* et assurer une autonomie collective	**Orientation des investisse-ments vers l'IA responsable pour le bien commun** Modèle de l'ISR (Investisse-ment socialement responsable). Investissement de l'État, des particuliers, en synergie avec les Fonds de Travailleurs	**Processus accéléré de mise à jour et de création de programmes professionnels** Avec cégeps, universités, ministères, ordres professionnels impactés par l'IA (ex. droit, santé)
Catégories d'instrument	Nouvelles formations	Nouvel acteur institutionnel	Nouveau mécanisme assurantiel	Incitatif	Dispositif de financement	Processus de planification

Comme dans le cas précédent sur le véhicule autonome, les propositions dénotent une véritable créativité institutionnelle (au-delà des exemples d'instruments très généraux donnés dans le livret du participant). Elles se situent également dans la lignée des enjeux identifiés à l'étape précédente, mais présentent aussi un enrichissement des idées. Si la littératie numérique est bien un objectif à l'agenda de la politique (ex. Stratégie numérique du Québec), c'est la nécessité qu'elle prenne de l'ampleur qui a été soulignée. Les autres mesures proposées sont inédites et invitent à concevoir de nouveaux dispositifs publics, multipartites ou collectifs pour assurer une véritable autonomie de la société québécoise face aux enjeux de l'IA dans le monde du travail. Dans ce sens, le groupe a pris le parti d'une responsabilité collective face à l'IA dans sa transition vers la société.

Troisième moment délibératif : PROPOSITIONS D'ENCADREMENT DE L'IA POUR 2018-2020

Ces mesures ont ensuite été mises en récit dans l'affiche. La « Une » du Journal de l'IA responsable du 18 février 2020 formulée par l'équipe était la suivante :

« PREMIÈRES MESURES DU COMITÉ INTERMINISTÉRIEL MIXTE SUR LA TRANSITION NUMÉRIQUE RESPONSABLE »

Le nouveau comité, créé le 14 mars 2018, suite à la journée de coconstruction pour la *Déclaration de Montréal*, s'est rapidement mis au travail et a élaboré une stratégie cohérente et intégrée avec toutes les parties prenantes. En ce début de 2020, le comité est fier d'annoncer le démarrage de 4 programmes :

1. Un nouveau fonds d'assurance numérique doté de 2 milliards (financé par les gains de productivité imputables à l'IA).
2. Une convention avec tous les cégeps et universités pour accélérer les renouvèlements de programmes de formation.
3. Un programme de soutien à la création de coopératives de travailleurs autonomes (contre la précarisation).
4. Un fonds de littératie doté de 10 milliards sur 5 ans sur la base d'un nouveau référentiel de compétences.

Cette « Une » du Journal, qui a été formulée à l'issue d'une discussion entre les participants, contribue à nouveau à la progression des idées. En effet, le Comité interministériel mixte sur la transition numérique responsable serait une création. Ce nouvel acteur institutionnel, issu d'une réflexion sur un scénario de 2025 sur l'impact de l'IA dans le monde du travail au Québec, pourrait représenter une nouvelle étape commune à plusieurs politiques publiques qui abordent bien la transition numérique et l'enjeu de la littératie numérique mais ne pose pas la question de l'impact social de l'IA : la **Stratégie numérique du Québec** du ministère de l'Économie, de la Science et de l'Innovation (MESI), la **Stratégie nationale sur la main-d'œuvre 2018-2023** du ministère du Travail, de l'Emploi et de la Solidarité sociale (MTESS), le **Plan stratégique 2017-2022** du ministère de l'Éducation et de l'Enseignement supérieur (MEES). Ce nouvel acteur, qui pourrait émaner d'une collaboration transversale entre la Commission des partenaires du marché du travail (CPMT), le Comité consultatif sur le numérique et la Commission mixte de l'enseignement supérieur, anticiperait en particulier les mutations du monde du travail et les nouveaux enjeux de formation et d'adaptation provoqués par le déploiement de l'IA dans les organisations publiques et privées au Québec.

6. PARTICIPANTS À LA COCONSTRUCTION ET ÉQUIPES DE TRAVAIL

Citoyens, professionnels et experts ayant participé aux ateliers et qui ont accepté de voir leur nom apparaître dans nos publications – Au Québec et à Paris

Sihem Neila Abtroon

Sébastien Adam

Béatrice Alain

Hassane Alami

Rana Alvabi

Alejandro Arreola-Alvarado

Gabriel Arruda

Jean-Claude Asssaker

Barthélémy Aucourt

Naomi Ayotte

Manon Babine

Maryluisa Barillas

Philippe Beauchemin

Stéphane Beaulieu

François Beauregard

Claude Bédard

Sylvain Bédard

Abdelkader Bekhti

Halim Benzaïd

Vincent Bergeron

Alexandre Berkesse

Karl Bherer

Emmanuel Bloch

Marise Bonenfant

Serge Bouchard

Caroline Boudreault

Lyne Bourbonnais

Véronique Boutier

Morgane Bravo

Robert Bruno

Beatrice Cassar

Ofelia Castaneda

Chantal Caux

Christian Chabot

Michel Chabot

Karine Charbonneau

François Charbonnier

Anne Chartier

Philippe Chartier

Guillaume Chicoisne

Pierre Choffet

Dominic Cliche

Lilen Colombino

Cristina Cotargasanu

François Côté

Jacques Coulombe

Lise Couturier

Alexis Cuglietta

Christian Cyr

Yvonne Da Silveira

Geneviève Dagneau

Hélène David

François-Michel De Rainville

David Décary-Hétu

Guillaume Déraps

Yves B. Desfossés

Michel Desy

Marc-Antoine Dilhac

Maxime Duban

Jean-Yves Dubé

Geneviève Dubois-Flynn

Mathieu Dubreuil-Cousineau

Geneviève Dufour

Arnaud Duhoux

Annie Dulude

Laurence Dumont

Mathieu Dumouchel

Benoit Dupont

Nicolas Dupras

Diane Duquette

Irina Entin

Julian Falardeau

Jacqueline Forien

Simon Frappier

Benoit Gagnon

Marie-Pierre Gagnon

Marina Gallet

Hortense Gallois

Sébastien Gambs

Véronique Gareau-Chiasson

Mathieu Gauthier-Pilote

Sylvie Gélinas

Thomas George

Gueno Gianni

Jean-François Gignac

Martin Gibert

Patricia Gingras

Béatrice Godard

Christian Goudreau

Gilles Gouin

Mervine Gowry

Alexandre Gravel

Michel Grou

Alexandre Guédon

Pascaline Guenou

Pierre Guillou

François Guité

Carl Hamilton

Simon-Pierre Harvey

Lucie Hébert

Ghiles Helli

Lucas Hubert

Aida Issa

Sabrina Jocelyn

Erwan Jonchères

Nico Julien

Debbie Jussome

Ed Khazen

Amy Khoury

Frederic Kleindienst

Andrée Labrie

Anne-Marie Lacombe

Marie-Claude Lagacé

Henri Lajeunesse

Karine Landry

Jean-Michel Lapointe

Jonathan Lasprilla

Sylvie Lavoie

Jean Latière

Louis Lecaer

Dominique Leclerc

Sarah Legendre Bilodeau

Pascale Lehoux

Claude Lejeune

Mélanie Levasseur

Elisabeth Limoges

Pamela Lirio

Robert Locas

Santiago Lopez

Aurélie Macé

Aicha Mafhoum

Suzanne Mainville

Mantas Manovas

Mathieu Marcotte

Jean-Pierre Marquis

Cloderic Mars

Marie Martel

Marième Mauger-Lavigne

Moussa Mekhnach

Natacha Mercure

Bruno Milia

Michael David Miller

Ann Mitchell

Erica Monteferrante

Farida Mostefaoui

Maria Moudfir

Jocelyne Mouton

Khalil Mouzawak

Vanessa Murray

Orly Nahmias

Vanessa Nantel

John Newhouse

Justin Ngoza

Zoonie Nguyen

Lisa Marlène Ntibayindusha

Catherine Olivier

Daniel Pascot

Florence Paulhiac

Ludovic Penet

Jorge Perez

Caroline Pernelle

Lorenzo Perozzi

Geneviève Perreault

Benoit Petit

Emmanuel Picavet

Louis Piette

Frédérick Plamondon

Pier-Luc Plante

Kamila Podgorska-Gilbert

Keith Poitras

Julie Politi

Philippe Polveche

Thomas Poulin

Denis Poussart

Emmanuelle Praine

Louis-Philippe Pratte

Mariel Ramos

Diane Raymond

Catherine Régis

Laurence Renault

Cassie Rhéaume

Toussaint Riendeau

Anne-Marie Robert

François-Xavier Robert

Louis-Nicolas Robert

Nicolas Roby

Stéphane Roche

Marie Roy

Sara Russo-Garrido

Laurence Sabourin

Iger Sadoune

Marie-Noëlle Saint-Pierre

James Sangster

Sylvie Saucier

Anton Selikhov

Jean-François Sénéchal

Eric Shannon

Danielle Sicotte

Chantale Simard

Julie Simard

Jean-Hébert Smith-Lacroix

Karima Smouk

Yanis Taleb

Isabelle Tanba

Christian Tanguay

Marc Tomkinson

Daniel Tremblay

Jérémy Trudel

Marie-Christiane Trudel

Félix Vaillancourt

Julie Verdy

Arnaud Vicari

Danael Villeneuve

Grant Wark

Bryn Williams-Jones

Lemy Wong

William Wong

Almina Yagoubi

Ming Yue

L’équipe de la coconstruction – Au Québec et à Paris

Simon Beaudoin-Gagnon, Maison des étudiants canadiens
Alexandre Beaudoin-Peña, Université de Montréal
Bhavish Beejan, Université Laval
Liam Bekirsky, Maison des étudiants canadiens
Karl Bherer, Université Laval
Alexis Bibeau, Université Laval
Pierre-Antoine Boutin-Panneton, Université Laval
Katie Charpentier-Bourque, Maison des étudiants canadiens
Arnaud Brubacher-Chouinard, Maison des étudiants canadiens
Dominic Cliche, Université Laval
Valentine Crosset, Université de Montréal
Rosemarie Desmarais, Maison des étudiants canadiens
Eve Gaumond, Université Laval
Martin Gibert, IVADO et Centre de recherche en éthique éthicien
Emilie Guiraud, Université Laval
Haykuhi Gutrez, Maison des étudiants canadien
Hubert Hamel-Lapointe, Université de Montréal
Audrey Houle, Université Laval
Samira Illourman, Maison des étudiants canadiens
Nico Julien, Université Laval
Henri Lajeunesse, Université Laval
Lauriane Long-Raymond, Maison des étudiants canadiens
Guillaume Macaux, Université Laval
Vincent Mai, Université de Montréal
Marième Mauger-Lavigne, Université de Montréal
Christophe Mondin, CIRANO
Orly Nahmias, citoyenne
Pauline Noiseau, Université de Montréal
Judith Paquet, Université Laval
Pierre-Luc Plante, Université Laval
Léa Ricard, Université de Montréal
Lynda Robitaille, Centre de recherche en données massives, Université Laval
Jason Stanley, Université de Montréal
Yanis Taleb, Université de Montréal
Clémence Varin, Université Laval
Nathalie Voarino, Université de Montréal
Camille Vézy, Université de Montréal
Alessia Zarzani, Université de Montréal

Les experts consultés

Sylvain Bédard, coordonnateur patient au Centre d'excellence sur le partenariat avec les patients et le public

Louise Béliveau, vice-rectrice aux affaires étudiantes et aux études de l'UdeM

Guillaume Chicoisne, directeur des programmes scientifiques, IVADO

David Décary-Hétu, professeur adjoint à l'École de criminologie de l'UdeM

Pierre-Luc Déziel, professeur, Faculté de droit de l'Université Laval, membre du CRDM

Thierry Karsenti, professeur titulaire à la Faculté des sciences de l'éducation de l'UdeM

Jihane Lamouri, coordonnatrice à la diversité, IVADO

Lyse Langlois, directrice de l'Institut d'éthique appliquée (IDÉA)

François Laviolette, professeur, Faculté des sciences et génie de l'Université Laval, directeur du Centre de recherche en données massives (CRDM)

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

Nicolas Merveille, professeur à l'École des sciences de la gestion de l'UQAM

Gregor Murray, directeur au Centre de recherche interuniversitaire sur la mondialisation et le travail, UdeM

Catherine Régis, professeure agrégée à la Faculté de droit de l'UdeM

Nicolas Roby, agent de recherche au Centre de recherche interuniversitaire sur la mondialisation et le travail, UdeM

Frank Scherrer, professeur titulaire à l'École d'urbanisme de l'UdeM

Marie-Odette St-Hilaire, architecte de solutions TI, Science de données, Service des technologies de l'information, Ville de Montréal

L'équipe de gestion

Isabelle Bayard, adjointe à la vice-rectrice à la recherche, à la découverte, à la création et à l'innovation

Joliane Grandmont-Benoit, développement numérique et coordonnatrice de projets, vice-rectorat aux affaires étudiantes et aux études

Anne-Marie Savoie, coordinatrice des travaux de la Déclaration, relations avec les partenaires et communications, vice-rectorat à la recherche, à la découverte, à la création et à l'innovation

L'équipe de recherche et analyse

Valentine Crosset, candidate au doctorat en criminologie, Université de Montréal

Jean-François Gagné, chercheur au CÉRIUM, Université de Montréal

Vincent Mai, doctorant en robotique, Mila, Université de Montréal

Mario Ionut Marosan, maîtrise en philosophie politique, Université de Montréal

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

Loubna Mekki-Berrada, doctorante en neuropsychologie, Université de Montréal

Christophe Mondin, professionnel de recherche chez CIRANO

Camille Vézy, doctorante en communication, Université de Montréal

Nathalie Voarino, coordonnatrice scientifique, candidate au doctorat en bioéthique, Université de Montréal

Alessia Zarzani, Ph.D en aménagement, Université de Montréal et Ph.D en Paysage et environnement, Université la Sapienza de Roma

L’équipe de coordination à Paris

Jacques-Henri Gagnon, chef, Communication, Jeunesse et relations universitaires, Ambassade du Canada en France

Hanane Hadjiloum, chargée des communication Maison des étudiants canadiens

Christine Métayer, directrice de la Maison des étudiants canadiens

Clément Thiébault, délégué commercial, Technologie de l’information et communication, Ambassade du Canada en France

Les partenaires ayant contribué à la coconstruction de l’automne

Les étudiants du Comité intersectoriel étudiant (CIÉ) des Fonds de recherche du Québec, participants aux Journées de la relève en recherche de l’ACFAS

Les professionnels membres de la Coalition de la diversité des expressions culturelles (CDEC-Canada)

Les élus et employés des différentes centrales syndicales ayant participé à la journée de réflexion sur l’IA, organisée par le Syndicat de la fonction publique et parapublique du Québec (SFPQ)

ANNEXES

ANNEXE 1 – **LES ATELIERS DE COCONSTRUCTION : DESCRIPTION DÉTAILLÉE ET FONCTIONNEMENT**

Les cafés citoyens

Les cafés citoyens sont des rencontres de 3 heures en bibliothèques publiques. Inclusives, ces rencontres sont ouvertes à tous les citoyens, et se déroulent de manière conviviale. Ces rencontres s'appuient sur l'esprit du café citoyen.

Le café citoyen (« world café ») est un dispositif de conversation convivial visant à faciliter le dialogue constructif et le partage d'idées. On recherche l'ambiance d'un café où les participants débattent d'une question en petits groupes. À intervalles réguliers, les participants changent de table. Un hôte reste à la table et résume la conversation précédente aux nouveaux arrivés. Les conversations en cours sont alors « fécondées » par les idées issues de la conversation précédente. Au terme du processus, les principales idées sont résumées lors d'une assemblée plénière, et les possibilités de suivi sont soumises à la discussion[30].

Cette méthode du café citoyen a notamment été adaptée et enrichie par plusieurs éléments :

- Une introduction sur la Déclaration de Montréal et les enjeux éthiques et sociaux de l'IA
- La lecture de scénarios prospectifs sectoriels en 2025 pour déclencher la discussion
- La contribution des participants à une seule table de discussion, pour permettre la délibération la plus approfondie possible
- L'utilisation d'une affiche pour documenter les discussions
- La distribution d'un cahier du participant présentant les principes de la Déclaration de Montréal IA responsable, un lexique, ainsi qu'une typologie exemplifiée des recommandations possibles

Ci-dessous, le déroulement type des cafés citoyens :

Tableau 5 : Déroulement type des cafés citoyens

Étapes	Heure	Description
Accueil	13 h à 13 h 30	Cafés et collations
Découverte de l'IA et de ses enjeux éthiques et sociaux	13 h 30 à 14 h	**Introduction pédagogique** Introduction aux enjeux éthiques et sociaux de l'intelligence artificielle (*Déclaration de Montréal IA responsable*), présentation des scénarios en 2025 et du déroulement de l'activité.
Café citoyen	14 h à 16 h	- Trois à quatre îlots thématiques (sur l'IA en santé, en justice, en éducation, dans la ville, dans le monde du travail) sont animés par un facilitateur. Chaque îlot accueille un petit groupe de participants (6 à 10) pour une discussion de 1 heure sur les enjeux éthiques et sociaux de l'IA à partir d'un scénario sur l'IA en 2025. - Les participants sont ensuite invités à imaginer en équipe la « Une » du Journal en 2020 (titre et premier paragraphe) sur une initiative importante à adopter au Québec pour un déploiement responsable de l'IA.
Synthèse en séance plénière	16 h à 16 h 30	**Synthèse des discussions en plénière** par les animateurs des affiches des différents îlots thématiques et discussion collective.

[30] Définition provenant de l'Institut du nouveau monde

Les journées de coconstruction

Ces rencontres d'une journée ont mobilisé des citoyens, parties prenantes et experts, pour à la fois approfondir les enjeux sectoriels de l'IA en société et formuler des recommandations. Ces journées se sont appuyées sur le modèle du codesign qui se situe au croisement du design, de la participation et de la prospective : la mobilisation de scénarios d'usages et de prototypes inconnus comme déclencheurs de discussions, moyen de défixation cognitive et véhicules d'exploration (c'est la dimension « design ») ; des dispositifs de participation collective mobilisant des acteurs issus de multiples horizons, citoyens, organismes comme experts (pour la dimension collective du « co »)[31].

L'approche prospective retenue pour les présents travaux consiste à se projeter dans un futur proche (2025) pour opérer un détour imaginaire et penser ensuite rétrospectivement des chemins innovants pour relier le présent aux futurs les plus souhaitables.

Michel De Certeau, dans son ouvrage *La culture au pluriel* souligne bien le jeu avec l'altérité de la prospective : « le futur interpelle le présent sur le mode de l'altérité »[32]. Et Georges Amar, dans un article sur la prospective conceptive insiste sur l'importance de la mise en récit de l'inconnu pour construire un futur ouvert : « Nous préférons du connu inefficace à un inconnu prometteur. La fonction de la prospective est de travailler l'inconnu, de lui donner des mots, des concepts, du langage. Afin que tout en demeurant inconnue, elle devienne plus abordable, qu'elle donne prise à la réflexion, à l'action »[33].

Ci-dessous, le déroulement type des journées de coconstruction :

Tableau 6 : Déroulement type des journées de coconstruction

Étapes	Heure	Description
Accueil	8 h 30 à 9 h	Café et viennoiseries
Mot de bienvenue et découverte de l'IA	9 h à 10 h	**Introduction :** principes de l'intelligence artificielle, les enjeux éthiques de l'IA (*Déclaration de Montréal*) et les scénarios prospectifs
Prospective en équipe	10 h à 11 h 30	**Prospective en équipe :** à partir d'un scénario déclencheur sectoriel et des principes de la *Déclaration de Montréal*, formuler les enjeux éthiques et sociaux en 2025 soulevés par le scénario et explorer comment une controverse éthique pourrait surgir ou s'amplifier
	11 h 30 à 12 h 30	**Plénière :** tour de table en plénière des enjeux éthiques et sociaux identifiés pour 2025 par chaque équipe et discussion avec l'ensemble des participants.
Repas sur place	12 h 30 à 13 h 30	Repas
Formulation de recommandations	13 h 30 à 14 h 45	**Formulation de recommandations** Travaux en équipe : à partir des enjeux éthiques identifiés le matin, formuler des recommandations (règlements, codes sectoriels, labels, politiques publiques, programmes de recherche, etc.) à mettre en place dès 2018-2020 au Québec.
	15 h à 16 h	**Exposés en plénière** des équipes et discussion collective
Conclusion de la journée et suites	16 h à 16 h 30	Retour et observations de la journée

[31] Méthode du codesign développée par le Lab Ville Prospective de l'UdeM, www.labvilleprospective.org

[32] Michel De Certeau, *La culture au pluriel*, p. 223, Paris, Seuil, 1993

[33] Georges Amar, Prospective conceptive : pour un futur ouvert, revue *Futuribles*, 2015, n. 404, p. 21

ANNEXE 2 – **LES SCÉNARIOS PROSPECTIFS DE LA COCONSTRUCTION DE L'HIVER**

ÉQUIPE DE RÉDACTION DES SCÉNARIOS

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 (CRÉ)

Valentine Crosset, candidate au doctorat en criminologie, Université de Montréal

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

Martin Gibert, conseiller en éthique pour IVADO et chercheur au Centre de recherche en éthique

Vincent Mai, doctorant en robotique, Université de Montréal

Christophe Mondin, professionnel de recherche chez CIRANO

N**athalie Voarino,** coordonnatrice scientifique, candidate au doctorat en bioéthique, Université de Montréal

Camille Vézy, doctorante en communication, Université de Montréal

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

Cette annexe présente les résumés de tous les scénarios sur l'IA utilisés dans cette première phase de coconstruction, et l'intégralité de cinq d'entre eux. Imaginés se déroulant en 2025, au Québec, ils furent à la base des débats et des délibérations sur les questions éthiques suscitées par l'intelligence artificielle. L'horizon de 2025 a été choisi pour se situer dans un avenir proche, au cœur de la décennie 2020-2030 qui devrait être celle du déploiement intensif de l'intelligence artificielle dans la société.

1. L'ensemble des scénarios résumés par thème

De février à mai 2018, dix-huit scénarios ont été mis en débat. Le tableau ci-dessous présente un résumé succinct de ces scénarios.

Tableau 7 : Résumé des scénarios

Thème	Scénario sur l'IA en 2025	Résumé du scénario sur l'IA en 2025 au Québec
1. Santé prédictive	Les jumeaux numériques en santé	Olivier apprend qu'un de ses 126 jumeaux numériques a reçu un diagnostic de dépression. Doit-il consulter ?
	Une assurance santé discriminante	L'assureur d'Olivier lui demande de changer de style de vie, sur la base de ses données personnelles. Peut-il refuser sans subir de conséquences ?
	Vigilo, un robot à domicile pour personnes âgées	Soline a 80 ans et elle vit à domicile avec Vigilo, son robot compagnon. Celui-ci rapporte régulièrement à la famille des diagnostics prédictifs sur sa santé. Soline souhaite-t-elle tout divulguer ?
	Une décision thérapeutique à l'hôpital	Un médecin expérimenté et un algorithme de reconnaissance médicale ne sont pas entièrement d'accord sur un diagnostic.
2. Ville intelligente	Voitures autonomes (réglage de l'algorithme et partage de la rue)	Pour assurer sa politique zéro accident, la Ville a mis en place des barrières de sécurité sur les axes où les véhicules autonomes peuvent aller à une vitesse « rapide » (50 km/h). Il s'ensuit une controverse sur le partage de la rue.
	Voitures autonomes (usage contingenté)	Les voitures autonomes sont devenues un service d'usage partagé pour les citadins. Des critères de priorité d'accès sont gérés par une IA dans le but de maximiser la croissance économique prédictive de la ville.
	Un frigo connecté qui vous veut du bien (*nudges*)	Une famille a acheté un frigo intelligent comportant un programme de *nudges* (« coups de pouce ») pour l'inciter à manger plus sainement et diminuer ses risques de maladie. Comment se partageront les gains de ce système entre l'assureur et la famille ?
	Une cote sociale basée sur l'empreinte carbone	La consommation d'une famille est encadrée et suivie de manière à prévenir un effet négatif sur l'environnement.
	Un jouet intelligent pas si fidèle que ça !	Jusqu'où peut aller la loyauté d'un jouet intelligent envers un enfant ? Est-ce la même que celle d'un ami ?

Thème	Scénario sur l'IA en 2025	Résumé du scénario sur l'IA en 2025 au Québec
3. Éducation prédictive	AlterEgo, IA d'aide à l'apprentissage scolaire	Une IA permet d'aider des élèves à apprendre plus efficacement, grâce à des devoirs et exercices personnalisés. L'enseignante a-t-elle toujours toute son autonomie professionnelle ?
	AlterEgo2, IA d'aide à l'orientation scolaire	Une IA oriente les élèves vers des métiers où la probabilité de réussir est très forte. Basé sur leur historique de données scolaires, le choix représentera-t-il vraiment le désir de l'élève ?
	Nao, une IA d'aide à la préparation de conférences	Une IA aide un conférencier à monter sa présentation et à l'actualiser en cours de conférence, au fil des réactions des étudiants.
4. Police prédictive et système judiciaire	Une arrestation préventive dans l'espace public	Le croisement des données personnelles d'Alexandre le classe depuis peu en individu potentiellement à risque. Suite à un comportement étrange dans l'espace public, il se fait arrêter de façon préventive.
	Une décision de libération conditionnelle	Un juge prend une décision d'ordonnance de probation pour une prévenue, contre la recommandation de l'algorithme. Celui-ci anticipe une probable récidive, mais sans tenir compte d'un nouveau programme de réinsertion (sans historique de données).
5. Monde du travail	Une IA pour optimiser l'ambiance au travail	Le département des ressources humaines d'une entreprise utilise une IA avec forage de données pour évaluer les styles de conduite de ses employés et les aligner sur sa norme de « bonne ambiance au travail ».
	Une cote sociale basée sur l'empreinte carbone	La consommation d'une famille est encadrée et suivie de manière à prévenir un effet négatif sur l'environnement.
	Une IA de recrutement comme passage obligé vers l'emploi	Tous les candidats à un emploi sont recrutés selon une vidéo analysée par IA, dans le but d'éliminer tout préjugé, favorable ou non. La neutralité dans le recrutement est-elle réelle et souhaitable ?
	Une restructuration socialement responsable	Une entreprise de logistique durable doit intégrer massivement l'IA dans plusieurs de ses services pour rester compétitive. Mais elle souhaite le faire de manière socialement responsable.
	Un nouveau comité sur la formation professionnelle	Le comité de la formation professionnelle d'une entreprise accueille de nouveaux membres : les représentants des robots collaborateurs. Tout le monde n'a pas le même point de vue sur cette évolution.

2. Cinq scénarios complets

Les cinq scénarios choisis explorent chacun une situation possible en 2025 pour un des thèmes abordés dans cette première phase de la coconstruction de la Déclaration de Montréal : la santé prédictive, l'éducation prédictive, la ville intelligente, la justice prédictive, et le thème transversal des mutations dans le monde du travail.

Chaque scénario présente le récit d'un cas qui a été construit en combinant plusieurs dimensions : une problématique sectorielle, une expérience d'usagers en 2025, un dispositif d'apprentissage mobilisant des données et une ou des techniques d'intelligence artificielle, et enfin, des enjeux éthiques et sociaux.

Tableau 8 : Constitution de cinq scénarios par thème

Scénarios sur l'IA en 2025	Jumeaux numériques	Voiture autonome	AlterEgo	Libération conditionnelle	Restructuration responsable
Thèmes	1. Santé prédictive	2. Ville intelligente	3. Éducation prédictive	4. Police prédictive et système judiciaire	5. Monde du travail
Problématique sectorielle	La santé préventive et personnalisée par profil similaire	Sécurité et partage de la rue	L'apprentissage personnalisé à l'école	La prise de décision du juge dans l'incertain	La gestion préventive et socialement responsable des mutations
Types d'apprentissage en IA	Partitionnement de données (clustering) en groupes homogènes par apprentissage non supervisé	Algorithmes des véhicules autonomes pour la vision, la prise de décision (apprentissage supervisé et par renforcement)	Apprentissage supervisé (concentration des élèves) et par renforcement (politiques de suites de devoirs)	Apprentissage supervisé sur les cas passés de récidive	Toutes les IA dès lors qu'elles impliquent des mutations dans les entreprises et les administrations
Enjeux éthiques et sociaux (exemples)	Vie privée : la confidentialité des données	Justice : le partage équitable de l'espace public	Vie privée : la confidentialité des données des élèves	Autonomie et connaissance critique dans la prise de décision	Justice : le partage équitable des gains de productivité

Thème 1 :
SANTÉ PRÉDICTIVE

Scénario de départ :
LES JUMEAUX NUMÉRIQUES

10 MARS 2025. Olivier reçoit une notification sur son téléphone lui indiquant qu'un de ses jumeaux numériques vient de recevoir un diagnostic de dépression.

Des jumeaux numériques sont des personnes qui partagent les mêmes caractéristiques biologiques et qui ont des profils de santé similaires. Toutes les données relatives à la santé d'Olivier sont collectées par Santé Canada depuis décembre 2023. Certaines proviennent de l'application santé de son téléphone – comme le nombre de pas qu'il effectue chaque jour ou ses heures de sommeil – et de ce qu'il partage publiquement sur les réseaux sociaux – données rachetées aux compagnies Alphabet et Baidu. Elles sont croisées avec les données qui proviennent directement du système de santé concernant son historique de maladies et ses prédispositions génétiques. Ces données sont mises en relation avec celles de l'ensemble de la population dans le « nuage de santé mondial », piloté depuis 2023 par l'Organisation mondiale de la Santé, qui permet de définir les profils de santé des individus, afin d'offrir à chacun une médecine de précision et une prévention ciblée et hautement personnalisée.

Olivier découvre donc ce matin-là qu'il est susceptible de développer la même pathologie qu'un de ses 126 jumeaux de santé numériques. Face à l'annonce de ce pronostic, l'algorithme de Santé Canada fait les recommandations suivantes à Olivier :

- Se rendre dans un centre spécialisé en santé mentale afin de recevoir un traitement préventif adapté ;
- Diminuer sa charge de travail à moins de 40 heures par semaine ;
- Augmenter son activité physique, en concordance avec les études sur les effets bénéfiques du sport sur la prévention de la dépression.

Olivier décide d'ignorer ces recommandations, car il travaille présentement à un contrat particulièrement déterminant pour sa carrière. Cependant, au cours des mois suivants, il apprend que 25 de ses jumeaux numériques ont reçu un diagnostic similaire.

Thème 2 :
VILLE INTELLIGENTE

Scénario de départ :
VOITURE AUTONOME – RÉGLAGE DE L'ALGORITHME ET PARTAGE DE LA RUE

AUTOMNE 2025. Les arrondissements du Plateau-Mont-Royal et de Rosemont–La Petite- Patrie se sont rejoints pour créer une zone pilote à Montréal où la circulation est organisée en priorité pour les véhicules électriques autonomes.

Les véhicules autonomes de particuliers ou en autopartage (Communauto, Car2go et les nouvelles capsules Goober) ainsi que des navettes autonomes de la STM y circulent à une vitesse de 25 km/h pour assurer un maximum de sécurité des usagers, des cyclistes et des piétons (politique « 0 accident » de la ville). Cette politique garantit une fluidité, sans embouteillages, avec des feux de signalisation rendus dynamiques grâce à un réseau de capteurs connectés. Tout ceci permet aux usagers d'envisager une activité dans leur véhicule sans être dérangés par des mouvements saccadés, par exemple, travailler, écrire, ou écouter de la musique. Les véhicules avec conducteurs doivent s'adapter à ces vitesses sous peine d'amendes dissuasives. Le nouveau Centre de régulation du trafic autonome (CRTA) autorise cependant une vitesse de 50 km/h aux heures de pointe du matin et du soir sur certains grands axes, comme l'avenue Papineau, la rue d'Iberville et le boulevard Saint-Joseph. Pour assurer la sécurité des piétons et éviter qu'ils ne traversent ces axes de manière improvisée, des barrières de sécurité ont également été installées en bordure de ces grands axes.

Samia, 30 ans, habite à Rosemont. Elle est massothérapeute, profondément tournée vers la relation d'aide et militante pour les droits des animaux. Elle vit avec son compagnon Robin, informaticien, et son chat Linus, 4 ans. Dès que c'est possible, elle laisse Linus librement aller dans la ville et peut le repérer en permanence grâce à son collier connecté. La vitesse très modérée des voitures autonomes la rassure pour son chat. De plus, elle apprécie que dans cette zone pilote de Montréal, les voitures soient réglées en mode « altruiste », c'est-à-dire qu'elles préservent les intérêts du plus grand nombre de personnes, même si c'est au détriment de la personne qui est dans la voiture.

Mais depuis l'été, un groupe de cyclistes est agacé de voir apparaître les nombreuses barrières de sécurité qui confisquent l'espace public pour les voitures autonomes. Depuis la fin août, ils font, pour protester, des « parades des vélos libres » sur les boulevards de l'arrondissement au nom du partage de la rue pour tous les modes de transports écologiques, sans hésiter à se jeter sous les roues des voitures autonomes, en sachant que leur « réglage altruiste » les préserve du danger. Mais ce matin d'octobre, Samia, dans sa voiture, ne sait pas que son mari Robin a modifié - par amour - le réglage de sa voiture pour la rendre « égoïste » : elle préserve désormais en priorité les intérêts de la conductrice en cas d'accident. Lorsque Laurène, une militante des vélos libres, franchit la barrière de sécurité et se jette devant sa voiture sur le boulevard Papineau, celle-ci ne réagit pas comme anticipé. Il se produit un accident qui blesse sévèrement Laurène, car les techniciens au CRTA n'ont pas non plus abaissé la vitesse de 50 à 10 km/h quand elle a franchi la barrière de sécurité. Samia est en état de choc.

Thème 3 : ÉDUCATION PRÉDICTIVE

Scénario de départ : ALTEREGO, IA D'AIDE À L'APPRENTISSAGE À L'ÉCOLE

28 AOÛT 2025. Carmen fait sa troisième rentrée à l'école Thérèse-Casgrain. Comme l'an dernier, elle a été assignée à une classe de 6e année. Elle est impatiente d'utiliser les nouveaux moyens pédagogiques que la Commission scolaire a mis en place dans cette école pilote pour améliorer l'accompagnement des élèves et pour personnaliser l'enseignement.

L'année dernière, Carmen avait repéré très tardivement les difficultés d'apprentissage de Samuel, un élève dont le manque d'attention, les bavardages et le comportement parfois agressif à l'égard de ses camarades perturbaient la classe. Carmen expliquait ses faibles notes par un possible trouble de déficit d'attention (TDA) et en avait parlé aux parents de Samuel. Cela ne s'était pas très bien passé.

Cette année, tout allait changer grâce à AlterEgo, une intelligence artificielle qui assiste les professeurs. AlterEgo mesure en temps réel le degré d'attention des élèves, il identifie ce qui fait obstacle à leur compréhension et détecte les enfants en difficulté. Le dispositif est très simple : grâce à des capteurs logés dans un bracelet électronique et aux tablettes connectées sur lesquelles travaillent les enfants, AlterEgo détecte le stress ressenti par les enfants et le relâchement de leur attention. Il est aussi capable d'analyser les variations de vitesse de lecture, afin d'identifier les enfants qui ont des problèmes de compréhension.

Aujourd'hui, Carmen remet aux élèves leur bracelet et répond aux questions des parents qui ont été invités à assister à la première heure de cours. Les parents ont d'abord été un peu surpris par le dispositif, mais ils semblent maintenant séduits par les prouesses d'AlterEgo. Les enfants, eux, jouent avec leur bracelet électronique et n'arrêtent pas de lui poser des questions sur leur tablette: « AlterEgo,

c'est qui ta chanteuse préférée ? ». AlterEgo se familiarise ainsi avec tous les élèves et commence à enregistrer les premières données.

Carmen explique que son assistant fait aussi des recommandations pédagogiques. Il peut suggérer de supprimer des parties du cours jugées inefficaces ou inadaptées à l'apprentissage ou recommander des suites d'exercices personnalisés pour chaque élève. À la fin de la journée, Carmen reprend les recommandations d'AlterEgo et étudie le profil de chaque élève pour prévoir une adaptation de son enseignement. Cela améliore considérablement le suivi des élèves. « Avec AlterEgo, fini le stress des examens ! » lance Carmen. Et c'est vrai : l'évaluation des élèves est désormais presque continue.
Elle s'empresse néanmoins de rassurer certains parents dubitatifs ; il y aura toujours des examens et l'évaluation continue n'est, pour l'instant, qu'une indication complémentaire. Un père demande à Carmen : « Qui corrige les examens ? C'est AlterEgo aussi ? » Carmen sourit et conclut sa présentation en plaisantant : « Quand je dois travailler le soir, c'est certain que j'aurais bien besoin d'AlterEgo pour s'occuper de mes enfants Lola et Emiliano. Ça viendra un jour ! »

Thème 4 :
JUSTICE ET POLICE PRÉDICTIVE

Scénario variante :
DÉCISION DE LIBÉRATION CONDITIONNELLE

AUTOMNE 2025. Sylvia, 29 ans, était en couple avec Jean depuis dix ans. Lorsqu'elle a appris que Jean l'avait trompée, elle a cherché à se venger en piratant son frigo connecté.

Connaissant l'allergie sévère de Jean aux arachides, son frigo, qui communiquait sa liste d'épicerie à un magasin partenaire, la formatait en fonction de cette caractéristique. Toutefois, lorsque Sylvia a piraté le système, l'allergie aux arachides de Jean n'apparaissait plus dans les paramètres par défaut et le frigo a produit une liste inadaptée à ses besoins de santé. En mangeant un plat préparé contenant une faible dose d'arachides, Jean a commencé à avoir des difficultés à respirer et a dû se rendre d'urgence à l'hôpital.

Sylvia a été arrêtée pour son délit. Au moment de son jugement, l'algorithme a calculé qu'il y avait 80 % de chances qu'elle récidive dans les deux années à venir, lui assignant une peine de deux ans de prison et une amende de 10 000 $.

Pour arriver à cette recommandation, l'algorithme a calculé le risque sur la base de plusieurs facteurs :

> Des facteurs historiques statiques, à savoir l'âge auquel Sylvia a commis sa première infraction et ses antécédents criminels (Sylvia avait déjà piraté la commande du pilulier de sa mère à 18 ans, et le réseau de caméras de vidéosurveillance de son quartier à 25 ans) ;

> Des facteurs de risque dynamiques : l'occupation de Sylvia, ses fréquentations, ses relations amoureuses et familiales, les remords exprimés par Sylvia, etc.

Puis l'algorithme a rapproché le cas de Sylvia à un grand nombre de cas similaires.

À la suite de cette décision émise par l'algorithme, le juge a eu le choix de suivre la recommandation de l'algorithme ou de donner une ordonnance de probation à Sylvia, avec la condition qu'elle suive le tout nouveau programme de réhabilitation pour délinquants mais sans historique de *data*, donc sans interprétation possible par l'algorithme.

Le juge, qui est favorable à l'innovation sociale, a choisi la seconde option. Le programme de réhabilitation prévoit pour Sylvia de suivre une évaluation et un contrôle individualisé régulier pendant une période de deux ans et demi, ainsi que de trouver un travail légal. Face à ses compétences en piratage, il est aussi demandé à Sylvia de mettre son savoir-faire à contribution dans le domaine de la cybersécurité.

Thème 5 :
MONDE DU TRAVAIL

Scénario de départ :
UNE RESTRUCTURATION SOCIALEMENT RESPONSABLE

15 JANVIER 2025. Créé en 2020 à Montréal, Zéro Carbone Logistique (ZCL) est un nouveau leader mondial de logistique durable et a connu, en cinq ans, une très forte croissance. L'entreprise emploie actuellement 3000 personnes à Montréal.

Dès son lancement, ZCL a souhaité inscrire ses objectifs environnementaux et sociaux dans sa convention d'actionnaires en adhérant au statut de B Corp. et en suivant les recommandations de la norme ISO 26000 sur la responsabilité sociale des entreprises. Cette politique a été bénéfique pour ZCL car plusieurs fonds syndicaux et fonds d'investissement socialement responsables ont rapidement investi dans l'entreprise, qui est devenue une *start-up* verte emblématique du Québec.

Toutefois, ZCL est une entreprise qui doit être rentable, et elle fait face à une concurrence très féroce sur les coûts des services : assurer une valeur environnementale ne suffira pas pour prospérer. Comme beaucoup d'entreprises, elle a donc fait un audit financier et le rapport préconise fortement un scénario radical pour la pérennité de l'entreprise : investir massivement dans l'IA et la robotisation de plusieurs tâches dès 2020. Cela inclut le calcul des bilans de carbone des trajets, les camions autonomes électriques, le tri des colis, le routage des dirigeables et des bateaux électriques et le suivi administratif des dossiers. Au total, 1000 emplois sur 3000 pourraient être supprimés, et 1000 autres devraient évoluer vers des formes de coopération entre humains et cobots ! Pour la direction de ZCL, il n'est pas question de faire cette évolution de façon brutale, et elle souhaite mettre en place une « restructuration socialement responsable », en préparant soigneusement les collaborateurs à de nouveaux métiers.

Nabila, une des fondatrices de ZCL, propose alors la solution suivante : créer, en partenariat avec un des géants du web, une plateforme de traitement des données massives utilisées par les applications en IA de la logistique. Jean-Raymond, représentant syndical de l'entreprise, est très inquiet : il souligne que ces entreprises fonctionnent avec des salariés sous-payés qui passent 15 heures par jour à coder des données pour l'entrainement des algorithmes, et que ce n'est pas une solution respectable pour les collègues. Il préfèrerait mettre en place une plateforme coopérative de traitement des données. « Il en existe en Californie et elles sont plus proche de nos valeurs ». Mais un gros acteur du web est prêt à investir tout de suite dans les données massives de la logistique durable et à créer, avec ZCL, une filiale à Montréal qui pourrait employer une grande partie des 1000 personnes. Le temps presse ; leurs investisseurs les incitent à choisir le partenariat immédiat qui est le plus sûr, même s'il aura très certainement un effet sur l'image de ZCL. Nabila et Jean-Raymond avaient pourtant évoqué ces enjeux à plusieurs reprises depuis 2023 en comité de direction. Ils auraient aimé pouvoir demander conseil plus tôt à un service public, mais ils ne savaient pas à qui s'adresser et maintenant, c'est trop tard.

< >
Déclaration de Montréal
IA responsable_
</ >

PARTIE 2

PORTRAIT 2018 DES RECOMMANDATIONS INTERNATIONALES EN ÉTHIQUE DE L'IA

TABLE DES MATIÈRES

RÉDACTION

CHRISTOPHE MONDIN, professionnel de recherche chez CIRANO

MARTIN GIBERT, conseiller en éthique pour IVADO et chercheur au Centrede recherche en éthique

GUILLAUME CHICOISNE, directeur des programmes scientifiques, IVADO

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

1. INTRODUCTION

En décembre 2016, Corinne Cath et ses collègues de l'université d'Oxford et du Alan Turing Institute publiaient une analyse comparée des politiques en matière d'intelligence artificielle (IA) émanant du Parlement européen, de la Chambre des communes britannique et de la Maison-Blanche[1] des États-Unis. Ils concluaient que ces trois rapports identifiaient correctement différents enjeux éthiques, sociaux et économiques, mais manquaient d'une stratégie à long terme pour le développement d'une « bonne IA ». Qu'en est-il aujourd'hui ? Comment différents organismes gouvernementaux et non gouvernementaux envisagent-ils les changements que l'IA va amener dans la société ?

On gardera en tête que plusieurs événements sont survenus depuis décembre 2016, des événements qui ont changé les attentes du public et des gouvernements à l'égard de l'IA et, plus généralement, des technologies de l'information. Les premiers accidents de voitures autonomes ont eu lieu. Les révélations sur les tentatives de manipulation des dernières élections présidentielles américaines via Facebook, ainsi que l'affaire Cambridge Analytica qui a éclaté en mars 2018, ont suscité de vives réactions et fait craindre pour la bonne santé des démocraties. De même, l'image de Google est sortie quelque peu ternie de ses velléités de collaboration avec l'armée américaine. On aura donc certainement une lecture plus juste des rapports analysés dans le présent document si on les resitue dans ce contexte – et cela vaut tout particulièrement pour la déclaration de principes éthiques publiée par Google en juin 2018.

1.1 MÉTHODE

Pour brosser un portrait rapide de la situation en 2018, nous avons analysé sept rapports et déclarations de principes publiés récemment. Les fiches techniques des documents retenus sont détaillées dans la troisième section de ce document. Nous y avons ajouté les fiches de rapports examinés, mais non retenus. Ce qui a guidé notre choix est d'abord la présence de recommandations de nature éthique. C'est loin d'être toujours le cas. En effet, de nombreuses réflexions prospectives sur le futur de l'IA s'inscrivent dans une perspective principalement économique : comment, par exemple, développer un écosystème favorable aux entreprises innovantes en IA, quel plan stratégique pour le développement de l'IA dans tel ou tel pays ? Nous avons donc mis de côté les rapports principalement économiques de même que les recommandations économiques dans les rapports retenus. Par ailleurs, nous n'avons pas retenu de rapports qui s'intéressaient exclusivement à un domaine particulier, comme l'éthique de la recherche en robotique ou la régulation des voitures autonomes. L'objectif était d'examiner des recommandations d'ordre général et comparables les unes aux autres.

Dans notre sélection, nous avons aussi cherché une certaine diversité afin d'avoir un spectre assez large pour une comparaison. Ainsi, deux rapports (Villani et la Commission nationale de l'informatique et des libertés (CNIL)) sont en français, les cinq autres en anglais. Un rapport émane d'une entreprise privée (Google), trois d'organisations non gouvernementales (Institute of Electrical and Electronics Engineers (IEEE), Asilomar et AI Now) et trois autres présentent les politiques officielles d'un pays (United Kingdom Royal Society (UKRS), Villani et CNIL). Certains rapports ont donc une visée globale quand d'autres sont plus locaux. Par ailleurs, certains rapports sont relativement concis (Asilomar,

[1] Cath, C., Wachter, S., Mittelstadt, B. et al. Sci Eng Ethics (2018) 24: 505. https://doi.org/10.1007/s11948-017-9901-7

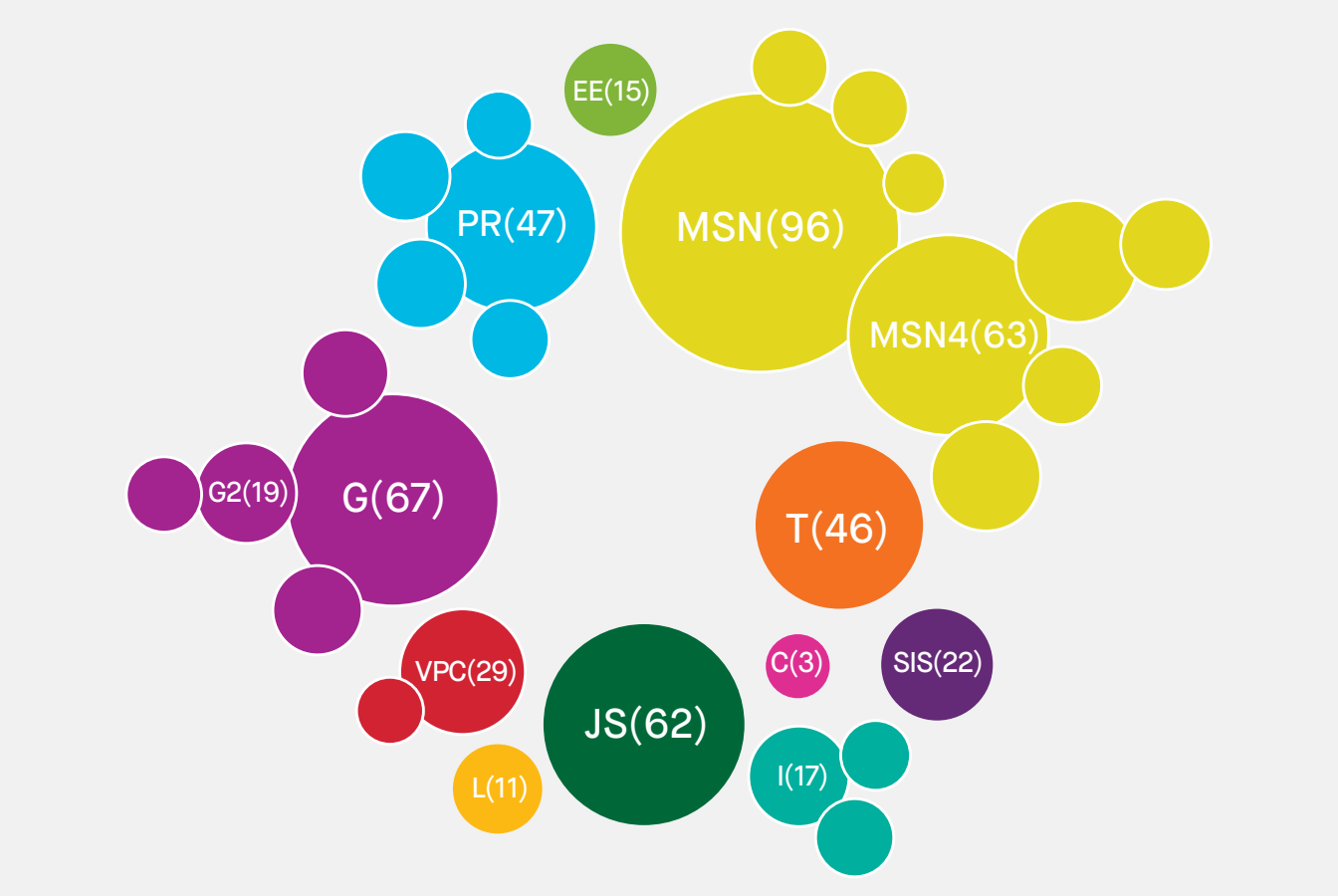
Google, AI Now), tandis que les autres sont beaucoup plus longs et développés, notamment parce qu'ils incluent des considérations d'ordre économiques.

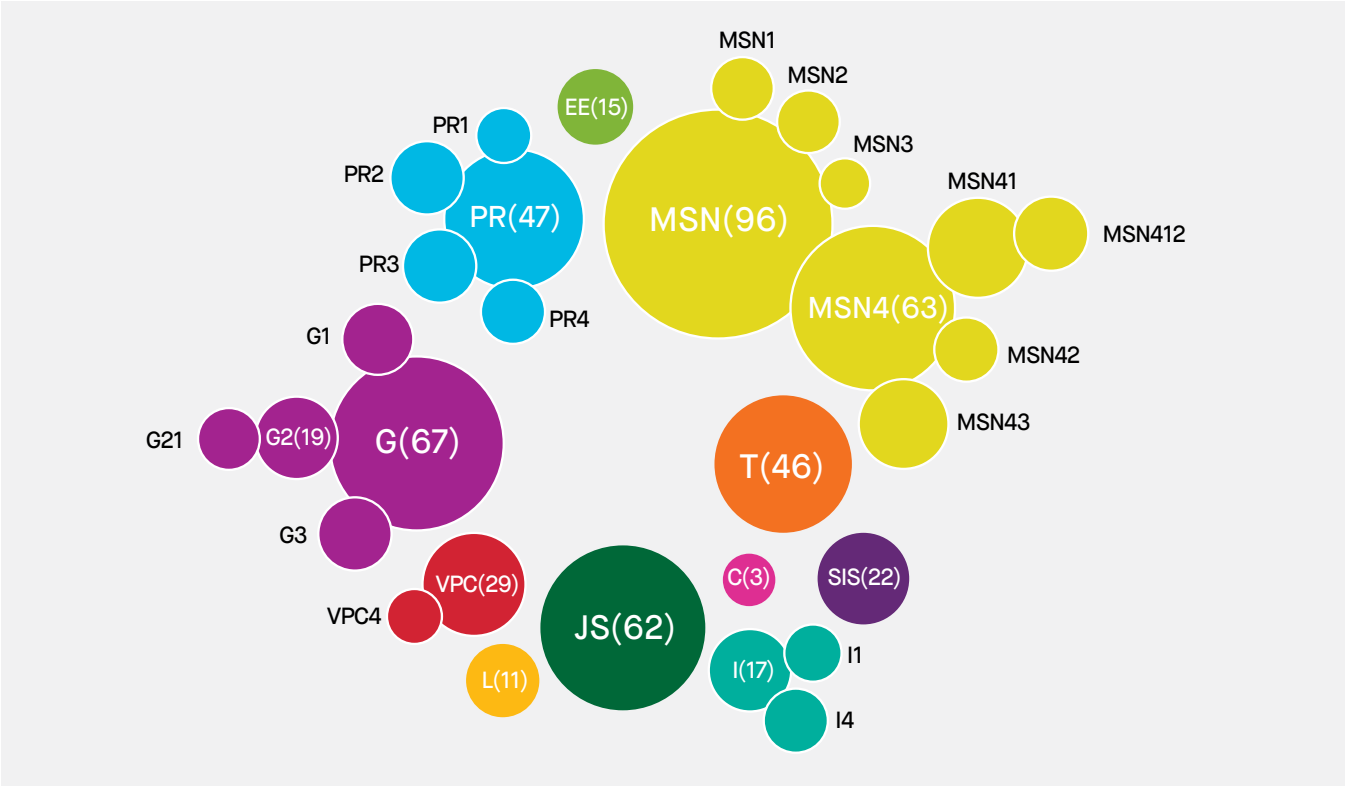
Dans les fiches techniques de la section 3, nous mettons aussi de l'avant la présence, ou non, de principes et de recommandations clairement identifiables. Nous appelons « principes » les propositions très générales, du type « l'IA devrait être bénéfique pour la société » tandis que les « recommandations » sont plus ciblées et relativement concrètes, du type « il faut développer des normes pour suivre la provenance et l'utilisation des jeux de données tout au long de leur cycle de vie ».

Du point de vue de la méthode, nous avons commencé par identifier dans les sept rapports les recommandations de nature éthique. Nous en avons retenu 230. Nous avons ensuite étiqueté ces recommandations selon sept catégories empruntées à la version préliminaire de la Déclaration de Montréal pour un développement responsable de l'IA, à savoir le bien-être, l'autonomie, la justice, la vie privée, la connaissance, la démocratie et la responsabilité – une même recommandation pouvant renvoyer à plusieurs catégories. L'avantage de ces étiquettes, c'est qu'elles renvoient d'emblée à ce qui nous intéresse, à savoir des valeurs morales. Bien sûr, étiqueter une recommandation relève souvent de l'interprétation et il n'est pas impossible que d'autres analystes soient parvenus à des résultats différents. Nous avons ensuite effectué des synthèses pour chaque valeur : ce sont elles qui sont présentées dans la deuxième section.

Afin d'éclairer les recommandations sous un jour différent, nous les avons également étiquetées à l'aide d'un ensemble raisonné de concepts clés. Ces concepts sont issus d'un index développé à partir des recommandations citoyennes établies lors de séances de réflexions collectives (dites de coconstruction) autour de la Déclaration de Montréal. C'est ainsi que nous avons obtenu le graphique ci-dessous.

Tableau 1 : Occurrence des concepts clés dans les sept documents examinés





LÉGENDE	DÉFINITION
C	Consentement
EE	Environnement et écologie
G	Gouvernance
G1	Collectivisme/Individualisme
G2	Gouvernance démocratique
G21	Communs numériques
G22	Participation citoyenne
G3	Gouvernance publique/privée
G31	Conflits d'intérêts
G32	Institutions publiques/ Compagnies privées
G33	Monopole
I	Influences
I1	Lobbyisme
I2	Manipulation
I3	Paternalisme
I4	Vulnérabilité des personnes
JS	Justice sociale
L	Libertés
MSN	Mutations socio-numériques
MSN1	Acceptabilité
MSN2	Transformation des activités
MSN3	Respect de l'humain
MSN4	Compétences

LÉGENDE	DÉFINITION
MSN41	Compétences de l'humain
MSN411	Dépendance à la technologie
MSN412	Littératie numérique
MSN413	Transformation des compétences humaines
MSN42	Complémentarité humain-IA
MSN43	Compétences de l'IA
PR	Partage de la responsabilité
PR1	Déresponsabilisation
PR2	Imputabilité
PR3	Responsabilité partagée
PR4	Souveraineté de la décision
SAA	Stress, alarmisme et angoisses
SIS	Sécurité et intégrité des systèmes
T	Transparence
VPC	Vie privée et confidentialité
VPC1	Anonymat
VPC2	Confidentialité
VPC3	Droit à l'oubli
VPC4	Propriété des données
VPC5	Intrusion

1.2

REMARQUES LIMINAIRES

Avant d'aller plus avant dans les fiches de présentation des rapports et les différentes synthèses par valeur, il nous semble utile de faire quelques remarques d'ordre général. Tout d'abord, on peut être frappé par la convergence des rapports : il est souvent difficile de déceler des contrastes saillants entre les recommandations des sept rapports. Cela s'explique sans doute en partie par une recherche de consensus : ces rapports visent à être rassembleurs et non polémiques ; ils évitent parfois les sujets potentiellement clivants en demeurant à un degré élevé de généralité. Mais il se pourrait aussi que cette convergence reflète tout simplement un accord de fond sur le type de relations qu'on devrait collectivement entretenir avec l'IA. Après tout, il n'est peut-être pas surprenant que tous s'accordent pour lutter contre l'automatisation algorithmique des discriminations ou pour promouvoir le renforcement du consentement dans la gestion des données des utilisateurs.

Il se pourrait aussi que cette convergence s'explique par le caractère assez homogène des sociétés dont émanent ces rapports : des pays occidentaux riches qui partagent globalement les mêmes valeurs démocratiques et libérales. À cet égard, on doit noter l'absence flagrante (l'éléphant dans la pièce) d'une question délicate : comment réguler l'IA sur le plan international ? En effet, les données, l'information et les algorithmes semblent tout particulièrement rétifs aux frontières territoriales. Ce que les autorités du Royaume-Uni, de France ou de tout autre pays peuvent accomplir demeurera ainsi toujours très limité en l'absence de coopération internationale. Mais est-ce véritablement envisageable ? De même, il ne faut pas oublier que les appels à lutter contre les discriminations et pour davantage d'égalité s'inscrivent dans un contexte global de croissance des inégalités. Autrement dit, les enjeux éthiques de l'IA peuvent difficilement être isolés des enjeux de justice internationale.

Cette convergence de fond dans les rapports examinés n'empêche pas pour autant ce qu'on pourrait qualifier de différences d'accents. Certains rapports mettent ainsi de l'avant des enjeux économiques et politiques (Villani et UKRS) quand d'autres s'en tiennent à des considérations juridiques ou éthiques. Par ailleurs, si tous se présentent comme des rapports d'experts, celui de la CNIL s'appuie en partie sur des consultations citoyennes. On doit aussi noter que la déclaration de principes de Google a une place à part dans la mesure où c'est la seule compagnie privée représentée parmi tous ces rapports. Cette déclaration est donc potentiellement porteuse de conflits d'intérêts, mais c'est aussi celle qui est la plus susceptible d'avoir des effets concrets internationaux étant donné le pouvoir de cette compagnie.

En ce qui concerne le contenu, le point de divergence le plus saillant est l'autorégulation des entreprises et le rôle des instances publiques dans la gouvernance des systèmes d'IA. Sans grande surprise, les rapports dont l'origine est gouvernementale, comme le rapport de la Royal Society britannique, le « UKRS », ou celui commandé par le gouvernement français, le « rapport Villani », proposent davantage de pistes de solutions émanant des institutions publiques. Ils préconisent aussi plus largement des outils législatifs pour répondre aux défis de l'avènement des systèmes d'IA — c'est aussi le point de vue de l'Institut des ingénieurs électriciens et électroniciens (IEEE). À l'inverse, les rapports d'AI Now et d'Asilomar abordent plutôt le problème dans la perspective des entreprises pouvant développer des outils de sécurité, des règles d'autorégulation et des guides de bonnes pratiques. Notons également que le rapport de la CNIL se distingue en proposant deux nouveaux principes, celui de vigilance et celui de loyauté des systèmes d'IA, tandis que le rapport Villani est celui qui accorde le plus de place aux enjeux environnementaux.

On remarquera enfin que ces rapports peuvent frapper par leur dimension pragmatique ou prosaïque. On est loin du lyrisme et des considérations existentielles qu'on trouve dans les ouvrages d'un Yuval Harari, d'un Nick Bostrom ou dans la littérature de science-fiction. L'accent n'est pas mis sur la rupture radicale qu'opère l'IA dans l'histoire de l'humanité, mais sur l'adaptation prudente et progressive aux innovations technologiques. De ce point de vue, on pourrait certainement réitérer le constat que Corinne Cath et ses collègues faisaient à partir des rapports de 2016 : la vision générale et à long terme d'une société avec une « bonne IA » demeure un chantier en cours.

2. SYNTHÈSE THÉMATIQUE DES RECOMMANDATIONS

Les sept rapports ou documents cités dans la prochaine section sont :

> **AI Now** : le rapport 2017 du AI Now Institute.

> **Asilomar** : les principes issus d'une conférence du Future of Life Institute.

> **CNIL** : le rapport de la Commission nationale (française) de l'informatique et des libertés.

> **Google** : les principes publiés par Google en juin 2018.

> **IEEE** : le rapport du Institute of Electrical and Electronics Engineers.

> **UKRS** : le rapport de la Royal Society britannique.

> **Villani** : le rapport « Donner un sens à l'intelligence artificielle » dirigé par le député français Cédric Villani.

BIEN-ÊTRE

Tous les rapports examinés comportent des recommandations qu'on peut explicitement associer au bien-être. Celles-ci sont d'ailleurs les plus nombreuses, ce qui n'est pas étonnant tant cette valeur est centrale et peut même dans une certaine mesure se confondre avec le bien. Les recommandations associées au bien-être renvoient notamment aux valeurs de compétences de l'IA, justice sociale, sécurité et intégrité des systèmes, vie privée et confidentialité, complémentarité humain-IA, et collectivisme/individualisme.

Il est possible de voir certaines tendances ressortir selon les rapports. AI Now insiste sur les enjeux de discriminations et de biais en demandant, par exemple, que les systèmes d'IA qui vont avoir un impact sur l'ensemble de la société soient développés par des personnes qui représentent la société dans toute sa diversité (AI Now, p.2). Villani lui emboitera le pas en précisant que tous les niveaux de la chaîne de conception de l'IA ont le devoir de représentativité de la société (Villani, p. 23). De son côté, la CNIL met l'accent sur la loyauté des algorithmes envers les personnes afin de ne pas les « trahir » en renforçant les discriminations (CNIL, p.48). L'IEEE met de l'avant la sécurité (IEEE, p.22) des systèmes d'IA qui devraient toujours être conçus de manière à profiter aux humains.

L'approche d'Asilomar se fait principalement sous l'angle de la recherche dont l'objectif devrait être de créer non pas une intelligence neutre, mais une intelligence bénéfique ; c'est pourquoi les financements devraient aller dans ce sens et inclure des disciplines comme les sciences sociales, l'éthique, le droit, la santé publique ou l'écologie. C'est aussi le cas pour UKRS qui demande au gouvernement d'encourager la recherche en développant des standards pour le partage de données (UKRS, p.8) et qu'on éduque les développeurs en apprentissage machine aux enjeux éthiques et sociaux (UKRS, p.9 et 12). On peut d'ailleurs dire que UKRS se distingue en mettant l'accent sur la recherche et l'enseignement.

De son côté, Villani, de pair avec de nombreuses considérations économiques, accorde une importance particulière aux effets de l'automatisation sur l'emploi. Il recommande par exemple de créer un « laboratoire public de la transformation du travail » et de « conduire un chantier législatif » (Villani, p.18) sur les conditions de travail à l'heure de l'automatisation. Ces recommandations s'inscrivent dans un projet plus large qui met de l'avant l'intérêt général et les enjeux de bien commun, notamment en santé : il faut développer l'IA pour la « détection précoce des pathologies, la médecine des 4P [prédictive, préventive, personnalisée et participative], la disparition des déserts médicaux, la mobilité urbaine à zéro émission » (Villani, p.15). Villani est aussi le seul rapport à mentionner comment favoriser la transition écologique (Villani, p.20), ce qui a bien évidemment des conséquences sur le bien-être.

Google, enfin, thématise la notion de bien-être dès son premier principe en affirmant que l'IA devrait

être socialement profitable. Les principes de la compagnie peuvent toutefois se distinguer des autres rapports dans la mesure où l'accent est mis sur la non-nuisance plutôt que sur la promotion du bien-être : il importe ainsi de faire des tests pour « éviter les risques de torts », de limiter les applications préjudiciables ou abusives, de ne pas développer des technologies potentiellement destructrices.

Mais du bien-être de qui parle-t-on dans ces rapports ? De façon plus ou moins explicite, il s'agit toujours du bien-être des humains : IEEE soutient par exemple qu'il faut donner la priorité au bien-être humain, en utilisant comme point de référence les meilleurs indicateurs de bien-être disponibles et largement acceptés (IEEE, p.25). Aucun rapport ne mentionne le bien-être animal. De même, les enjeux environnementaux lorsqu'ils sont soulevés (Villani, p.20) le sont dans une perspective anthropocentriste (par opposition à des perspectives pathocentriste, biocentriste ou écocentriste). La porte n'est pas pour autant fermée pour le bien-être des non-humains. En effet, l'idée d'alignement de l'IA avec les valeurs humaines, qu'on trouve par exemple dans Asilomar, laisse ouverte l'option de voir la compassion envers les plus vulnérables ou le souci des autres espèces comme une valeur humaine.

S'il est vrai que seul le bien-être humain est considéré, en revanche, on peut dire que les rapports sont « universalistes » dans la mesure où ils ne font pas de distinctions entre les sous-catégories de la population humaine – autrement dit, il s'agit de respecter l'universalité des droits humains. Par exemple, aucun rapport n'affirme que seuls une oligarchie, un État ou une organisation devraient en bénéficier – bien au contraire précise Asilomar. Autrement dit, les opportunités liées à l'avènement de l'IA doivent bénéficier à tous souligne Villani (Villani, p.23) qui note en même temps qu'il faut anticiper les impacts des changements technologiques, « en particulier pour protéger les populations qui sont déjà les plus fragiles » (Villani, p.18).

Quand ils évoquent le sujet, les rapports sont prudents quant à savoir qui devrait profiter de la richesse créée par l'IA (une question que les philosophes politiques désignent sous le nom de justice distributive). Ils en appellent surtout à la réflexion. Villani recommande ainsi d'instaurer « un dialogue social autour du partage de la valeur ajoutée » (Villani, p.19) tandis que UKRS préconise que la société prenne en compte de façon urgente la manière dont « les bénéfices de l'apprentissage automatique peuvent être partagés dans la société » (UKRS, p.12). Ce « temps de la réflexion » sur la redistribution des richesses trouve peut-être un écho dans l'appel assez courant parmi tous les rapports à enrichir la recherche en IA de collaborations avec les sciences sociales ou l'éthique (p. ex. Asilomar).

De son côté, la version préliminaire de la Déclaration de Montréal propose comme principe : « Le développement de l'IA devrait ultimement viser le bien-être de tous les êtres sentients ». Elle se positionne donc comme plus inclusive en assumant une perspective pathocentriste. On peut même dire que c'est un des éléments les plus originaux de la Déclaration de Montréal : ne pas considérer seulement le sort des êtres humains, mais celui de tous les individus qui pourraient être affectés par le développement de l'IA.

AUTONOMIE

On trouve des recommandations explicitement liées à la notion d'autonomie dans tous les rapports examinés – à l'exception de AI Now. Celles-ci sont tout particulièrement associées aux enjeux de compétences de l'humain, complémentarité humain-IA, compétences de l'IA, acceptabilité, vulnérabilité des personnes et justice sociale.

D'un point de vue général, c'est l'idée que l'IA doit respecter l'autonomie des êtres humains qui est défendue dans les divers rapports. Asilomar soutient par exemple que les systèmes d'IA doivent être fabriqués et opérés de manière à être compatibles avec les idéaux de dignité humaine, le respect des droits, des libertés et de la diversité culturelle. La CNIL (CNIL, p.57) va peut-être un peu plus loin puisqu'il s'agit non seulement de respecter l'autonomie mais de la promouvoir et ce, dès la phase de conception ou de design. Cette distinction entre

respecter et promouvoir renvoie en général, chez les philosophes, à celle entre une logique déontologique de respect des normes (l'autonomie comme droit) et une logique conséquentialiste de promotion des valeurs (l'autonomie comme bien). Toutefois, on se gardera de sur-interpréter ici le choix des termes. La CNIL précise même qu'il s'agit de corriger une situation puisqu'elle insiste sur l'importance de « remédier aux situations d'asymétrie », étant entendu qu'il ne peut y avoir d'autonomie véritable dans une situation où l'un des acteurs possède tout le pouvoir ou toute l'information. Pour la CNIL, promouvoir l'autonomie passe d'ailleurs par la sensibilisation des professionnels qui utilisent l'IA (CNIL, p.55).

Ce respect ou cette promotion de l'autonomie des utilisateurs s'exprime aussi avec l'idée que l'IA doit demeurer un outil, un instrument au service des utilisateurs ou, plus largement, des êtres humains. L'IEEE mentionne que les systèmes d'IA devraient toujours être subordonnés au jugement et au contrôle humain (IEEE, p.23). Cette idée fait écho au principe de Google pour qui les technologies de l'IA « doivent être soumises à une direction et contrôle humain approprié » (Google). Le rapport de la CNIL est d'ailleurs titré « Comment permettre à l'homme [sic] de garder la main ».

On peut voir cette quête d'autonomie comme le résultat d'un effort conjoint des entreprises qui fournissent l'IA et de ceux qui l'utilisent. Pour Asilomar, ce sont les humains qui doivent décider de la nécessité et de la façon de déléguer des décisions aux systèmes d'IA afin d'accomplir des objectifs choisis par des humains. La CNIL (CNIL, p.57) est plus concrète en notant que les utilisateurs devraient pouvoir « jouer » dans les paramètres d'un système donné, ce qui a notamment l'avantage d'en favoriser la compréhension. Pour Google, c'est aussi en termes d'information et de consentement que les compagnies doivent mettre l'IA au service des utilisateurs, en particulier « en fournissant une transparence et un contrôle approprié sur l'utilisation des données », ce qui nous rappelle que les enjeux d'autonomie et de vie privée ne sont jamais bien loin.

Une autre option semble être de sortir du paradigme de l'outil pour favoriser la complémentarité humain-machine non aliénante. Pour Villani (Villani, p.18) cette complémentarité pourrait s'appuyer sur le développement des capacités proprement humaines comme la créativité, la dextérité manuelle ou la capacité de résolution de problèmes. De nouveaux moyens semblent requis pour atteindre ce type d'objectifs : il faut de nouvelles médiations (Villani, p.23) ou une formation à la littératie numérique, dès l'école primaire et jusqu'à l'université pour tous les citoyens (CNIL, p.54).

La CNIL (CNIL, p.48) propose un principe de loyauté qui résume assez bien l'esprit de ce que pourrait être une bonne gestion de l'autonomie à l'ère de l'IA. « Un algorithme loyal ne devrait pas avoir pour effet de susciter, de reproduire ou de renforcer quelques discriminations que ce soit, fût-ce à l'insu de ses concepteurs ». Et cette loyauté doit se comprendre non seulement à l'égard des utilisateurs individuels que de la collectivité dans son ensemble – parce que c'est toute la société qui pourrait être affectée par des « décisions » algorithmiques non voulues explicitement. On y voit aussi comment les enjeux d'autonomie sont souvent adjacents à ceux de justice.

De son côté, la version préliminaire de la Déclaration de Montréal propose comme principe : « Le développement de l'IA devrait favoriser l'autonomie de tous les êtres humains et contrôler, de manière responsable, celle des systèmes informatiques. » En raison de son caractère très général, ce principe apparaît être en phase avec les différents rapports. Il s'en distingue légèrement en évoquant, dans sa formulation, l'autonomie des systèmes informatiques – là où les autres rapports semblent davantage se focaliser sur l'autonomie humaine et les risques qu'elle s'amenuise.

JUSTICE

On trouve des recommandations dans tous les rapports et les thématiques qui ressortent le plus sont : justice sociale, compétences de l'humain, complémentarité humain-IA, compétences de l'IA, et respect de l'humain.

L'idée principale est que l'intelligence artificielle, et les systèmes qui en utilisent le pouvoir, doivent conduire à une société plus juste, plus égalitaire (AI Now, p.2). Cette idée s'articule autour de deux principes :

1. l'IA doit avoir comme but de gommer les défauts de la société dans ces domaines (UKRS, p.12) ;
2. il faut prendre garde, en particulier lors des étapes de développement et de déploiement, à ne pas créer ou ne pas faire perdurer des injustices (Google).

Ces deux objectifs seront atteints en proposant des solutions à plusieurs niveaux.

Les avancées de l'IA doivent bénéficier à tout un chacun (Google). C'est l'idée de ruissellement (Villani, p.19) : les bénéfices (en service) et les richesses (en savoir-faire, en technique/technologie, en données accumulées) ne doivent pas être l'apanage des grandes entreprises privées (Villani, p.14) ou des strates supérieures de la société — qui peuvent être aussi bien la majorité de la population en termes de culture, religion, ou ethnie, qu'une minorité de la population en termes de revenus comme le « 1 % ». (Villani, p.22).

Les avancées de l'IA doivent viser un monde meilleur où les inégalités existantes sont prises en compte et combattues, dans le système judiciaire (Asilomar), dans l'attribution des soins en santé, ou en protégeant les populations habituellement laissées pour compte (AI Now, p.1 et 2 ; Villani, p.18 ; Google). Il faudrait par exemple créer une base de données nationale permettant d'objectiver les inégalités entre les femmes et les hommes au travail (Villani, p.23) afin de résoudre les problèmes de discrimination liés au genre. De même, il faut canaliser le développement de l'IA vers des applications qui contribuent à améliorer autant la performance économique que le bien commun.

Pour bénéficier à tout un chacun, l'IA doit être inclusive, et ceci à tous les niveaux (Villani, p.23). Cela signifie que dans toutes les étapes, de sa conception jusqu'à son déploiement et durant sa maintenance, un système d'IA devrait être examiné par des représentants de la société. Il importe de proposer des incitatifs pour inclure davantage les populations comme les femmes ou les minorités.

Des formations complémentaires en sciences sociales, en éthique, peuvent venir aider les concepteurs en leur faisant prendre conscience de ces enjeux et en leur fournissant les outils conceptuels et intellectuels pour y faire face (AI Now, p.1 et 2). De même, il faut encourager et supporter financièrement la recherche sur l'interprétabilité des algorithmes, leur robustesse, les questions d'égalité, de vie privée et de causalité (UKRS, p.13).

Enfin, la justice concerne aussi les institutions judiciaires qui peuvent être directement touchées par le développement de l'IA. Voici ce que proposent différents rapports :

> Il importe de développer un cadre légal pour garantir la justice sociale, la représentativité de tous dans la conception et l'utilisation des algorithmes, gommer les inégalités, et prévenir des abus ou déviances pouvant survenir avec une utilisation de l'IA non régulée (Asilomar).

> Il est nécessaire de faire une importante mise à jour de l'appareil judiciaire sur toutes les questions touchant à l'intelligence artificielle et à la donnée, en particulier sur les questions de souveraineté, de propriété, de citoyenneté de la donnée et de gouvernance (UKRS, p.12 ; Asilomar ; IEEE, p.22). De même, il faut conduire une importante réflexion sur la notion de transparence et ses critères d'évaluation si l'on veut juger de la conformité des entreprises utilisant des systèmes d'IA (IEEE, p.30).

> Ces cadres légaux et éthiques devraient être conçus en faisant appel à tous les acteurs de la société : la communauté scientifique, les pouvoirs publics, les industriels, les entrepreneurs et les organisations de la société civile (Villani, p.21). Les systèmes de contrôle devraient régulièrement être évalués pour s'assurer qu'ils remplissent correctement leur mission.

> De la même manière qu'il a été décidé qu'une entreprise est une entité juridique à part entière, il faut lancer une réflexion sur la nature juridique de l'IA elle-même (Asilomar).

> Lorsqu'une intelligence artificielle prend part à des décisions de justice, il faut mettre en place des mesures d'audit, d'interprétation, de vérification, et d'explication (Asilomar).

De son côté, la version préliminaire de la Déclaration de Montréal propose le principe suivant : « Le développement de l'IA devrait promouvoir la justice et viser à éliminer les discriminations, notamment celles liées au genre, à l'âge, aux capacités mentales et physiques, à l'orientation sexuelle, aux origines ethniques et sociales et aux croyances religieuses ». Avec cet énoncé, elle touche principalement à la justice sociale et aux problématiques d'égalité et d'équité, que cela soit en venant réparer les discriminations passées ou en anticipant les discriminations futures. La Déclaration de Montréal n'entre pas dans le détail des moyens d'atteindre ces objectifs, à l'inverse de plusieurs rapports qui suggèrent, par exemple, plus d'inclusion et de représentativité sociales dès l'étape de la conception des systèmes d'intelligence artificielle. Par ailleurs, elle n'aborde pas les implications spécifiques au monde judiciaire.

VIE PRIVÉE

Les recommandations portant explicitement sur la vie privée (*privacy,* en anglais) sont présentes dans tous les rapports considérés, AI Now excepté. Celles-ci sont notamment associées à des enjeux de vie privée et confidentialité, collectivisme/ individualisme, communs numériques, gouvernance, justice sociale, transparence, et sécurité et intégrité des systèmes.

À un niveau très général, la question de la vie privée se traduit par l'idée que l'utilisateur devrait avoir le contrôle sur ses données – on peut donc y voir un lien avec les enjeux d'autonomie. Asilomar soutient par exemple que les gens devraient avoir le droit d'accéder, de gérer et de contrôler les données qu'ils génèrent tandis que Google affirme que la protection de la vie privée devrait jouer un rôle important dans la conception des principes d'IA et dans le développement des systèmes d'IA. On notera toutefois que les rapports sont plutôt avares de principes généraux sur la vie privée. Tout se passe comme si la question était difficile à traiter à un tel degré de généralité.

La protection de la vie privée suppose des cadres de gouvernance divers, notamment des organismes de réglementations et d'autres qui fixent des standards (IEEE, p.22). Pour la CNIL (CNIL, p.45) c'est à la loi d'encadrer l'utilisation des données personnelles utilisées par l'IA. Un bon exemple est fourni par Villani (Villani, p.14) qui, à la suite du règlement général sur la protection des données (RGPD) européen, fait mention du droit à la portabilité, à savoir celui pour un utilisateur de récupérer les données qu'il a générées sur une plateforme pour les utiliser sur une autre plateforme.

Il peut être possible de distinguer deux tendances quant aux modèles socio-politiques qui déterminent la gouvernance des données. En effet, Villani et la CNIL semblent davantage favoriser une logique de la donnée comme bien commun, quand UKRS semble s'inscrire dans une logique plus « libérale » ou, à tout le moins, davantage centrée sur l'individu. Encore une fois, on se gardera de trop contraster ces approches, tant il est délicat de déduire une tendance générale à partir de quelques recommandations. Toujours est-il que Villani (Villani, p.14) plaide pour que la puissance publique impose « l'ouverture s'agissant de certaines données d'intérêt général ». On peut penser à des données médicales qui, mises en commun pourraient faire progresser la recherche et bénéficier à toute une population, ou à des données environnementales, par exemple, qui aideraient à lutter collectivement contre les changements climatiques. Cette proposition s'inscrit dans la continuité de la CNIL (CNIL, p.59), un autre rapport français qui propose que l'État se lance dans « un grand projet de recherche fondé sur des données issues de la contribution de citoyens exerçant leur droit à la portabilité auprès des acteurs privés. »

Pour UKRS, c'est plutôt l'importance de la protection de la vie privée dans la recherche scientifique qui est mise de l'avant. Il s'agit de protéger les individus. Ainsi, les chercheurs devraient tenir compte des utilisations futures potentielles des données qu'ils recueillent et intégrer cette dimension dans le consentement des participants à la recherche (UKRS, p.8). Ce souci doit être présent depuis la collecte des données jusqu'à son éventuel partage ou redistribution. Le contraste entre les deux logiques demeure toutefois assez artificiel dans la mesure où la CNIL propose, elle aussi, de développer des infrastructures de recherches « respectueuses des données personnelles » (CNIL, p.59), tandis que UKRS n'est pas opposé à la logique d'un « bien commun des données » lorsque celles-ci émanent de recherches financées par des fonds publics ou par des organismes de charités (UKRS, p.8).

On notera pour finir qu'il existe bien sûr un lien entre les enjeux de protection de la vie privée et ceux de justice puisque les données personnelles pourraient servir de base à des politiques discriminatoires. Cette dimension est présente dans la plupart des rapports.

De son côté, la version préliminaire de la Déclaration de Montréal, propose comme principe : « Le développement de l'IA devrait offrir des garanties sur le respect de la vie privée et permettre aux personnes qui l'utilisent d'accéder à leurs données personnelles ainsi qu'aux types d'informations que mobilise un algorithme. » Si l'on peut reconnaître à ce principe le mérite de proposer une synthèse plutôt en phase avec ce qui ressort des autres rapports, force est de constater qu'il n'épuise pas le sujet complexe et ramifié de la vie privée. En particulier, ce principe de la Déclaration de Montréal ne mentionne pas les enjeux de transparence qui, à bien des égards, sont le corollaire de la vie privée – et qui sont analysés dans la prochaine section, sur la connaissance.

CONNAISSANCE

On trouve des recommandations liées à la connaissance dans tous les rapports, et les thématiques qui ressortent le plus sont : justice sociale, transparence, compétences de l'humain, et littératie numérique.

Les deux principaux axes de réflexion sont le développement de la connaissance du public et celui des autorités qui vont valider ou vérifier les systèmes d'IA. En effet, l'autonomie du public et des organes de gouvernance, de même que la transparence, ne peuvent exister que si l'on offre la possibilité au public et au gouvernement de l'exercer, en leur fournissant d'un côté les mécanismes et les infrastructures nécessaires, et de l'autre, les formations, l'éducation et l'esprit critique.

Pour aiguiser l'esprit critique et la compréhension de ces nouvelles technologies, il faut instaurer une nouvelle littératie numérique (CNIL, p.54), dès la petite école et jusqu'à l'université, pour tous les citoyens. Il s'agit de promouvoir une nouvelle conception de l'autonomie intellectuelle et de la réflexivité des personnes vis-à-vis de l'ensemble des problématiques quotidiennes liées à l'IA (CNIL, p.57) – par exemple, comprendre ce que signifie donner son consentement. Autrement dit, il convient de remédier à des situations d'asymétrie entre les prestataires de services utilisant de l'IA et l'utilisateur/citoyen.

Afin de protéger le public, il est crucial d'éveiller sa conscience aux possibilités d'utilisation pernicieuse des systèmes d'IA. Cela suppose d'instaurer les bases d'une méthode éducative et des outils de mesure adéquats (IEEE, p.31), par exemple, un test de validation à l'école. Pour compléter cet apprentissage, encore une fois des notions d'éthique et de sciences sociales sont suggérées (IEEE, p.31). Les personnes les plus « à risque », c'est-à-dire celles identifiées comme étant davantage crédules et/ou celles pouvant subir de plus grandes conséquences de ces utilisations abusives, sont à cibler en priorité (IEEE, p.31).

En plus de celles destinées au grand public, de nombreuses recommandations s'adressent aux agents gouvernementaux, aux représentants élus qui vont voter les lois, au système judiciaire qui

va les appliquer et aux institutions qui en seront garantes (IEEE, p.31). D'autres secteurs « à risque », comme la médecine, les ressources humaines (recrutement) ou encore, le marketing devront être tout particulièrement vigilants (CNIL, p.55).

Évidemment, les concepteurs d'algorithmes et de systèmes IA sont eux aussi concernés par ces mesures : il est conseillé de compléter leur formation avec des sciences humaines dans le but de saisir les enjeux sociaux et économiques des solutions qu'ils conçoivent et de prendre conscience de l'impact que leurs solutions pourraient avoir en pratique (CNIL, p.55). Renforcer la diversité culturelle, sociale et de genre est une recommandation présente dans plusieurs rapports ; elle implique l'idée qu'en multipliant les représentants de la société à chaque étape de conception de l'IA, il devient possible d'avoir une meilleure connaissance de tous les paramètres, contextes, et points de vue à prendre en compte (CNIL, p.55).

Enfin, de nombreuses recommandations ciblent la connaissance nécessaire au bon fonctionnement des infrastructures de contrôle et d'évaluation des systèmes d'IA. Il faut d'abord instaurer des standards et des organes de régulations pour surveiller les différentes étapes du processus de conception des systèmes d'IA, et s'assurer qu'ils respectent les droits humains, les libertés, la dignité, la vie privée et la traçabilité (IEEE, p.22). Ces standards doivent être mis en place par des institutions publiques (IEEE, p.30) qui développeront des outils de mesure transparents, ouverts au public (AI Now, p.1), construits par des experts et des professionnels impartiaux.

La transparence des instances régulatrices est une recommandation qui apparaît régulièrement dans les rapports. L'ouverture au public de toutes ces méthodes d'évaluation lui permettra d'exercer et de faire valoir les connaissances acquises. Avec des utilisateurs formés et stimulés, avec des systèmes d'IA encadrés par des comités transparents et documentés, une dernière étape semble être de laisser les citoyens libres d'expérimenter, de déployer leur littératie numérique et d'exercer leur esprit critique. Il est par exemple suggéré que les diverses plateformes d'utilisation de systèmes d'IA fournissent de l'information sur la logique de fonctionnement de leurs algorithmes (CNIL, p.45 et 48). Des informations précises sur les données utilisées et la logique des algorithmes pourraient être disponibles sur les pages de profil des utilisateurs (CNIL, p.56). Pour favoriser la compréhension, ces derniers devraient pouvoir « jouer » avec les systèmes en faisant varier les paramètres (CNIL, p.57).

Dernier point : il faut assurer la transition en vérifiant et en améliorant les formations à l'école. La littératie numérique est définie de différentes manières, depuis l'éthique et l'esprit critique déjà mentionnés, jusqu'à la connaissance des principes clés de la programmation ou de l'apprentissage machine (UKRS, p.9). Il s'agit encore une fois de venir rééquilibrer les asymétries qui peuvent exister entre les utilisateurs, les développeurs et les citoyens. Pour contribuer à cette littératie numérique, il convient d'impliquer à la fois les gouvernements, les spécialistes en mathématiques et en programmation, les entreprises et les professionnels de l'éducation, dans le but d'insuffler de nouvelles connaissances nécessaires et suffisantes (UKRS, p.9). Plusieurs recommandations soulignent l'importance et l'intérêt d'inculquer des notions d'éthique, de sciences sociales et, aussi, de santé publique dans les activités pédagogiques (UKRS, p. 9).

Quant au système éducatif, il devrait aussi avoir pour mission de former une nouvelle génération de travailleurs et de chercheurs ayant les compétences nécessaires pour naviguer dans un monde imprégné de systèmes d'IA. Il s'agit non seulement de repenser la formation initiale à l'université, mais aussi les formations continues afin de proposer de nouvelles compétences aux travailleurs dont les tâches vont être profondément modifiées. De telles recommandations prennent tout leur sens dans un contexte de menace sur l'emploi dû au remplacement de l'humain par la machine (UKRS, p.9). Ce sont à la fois l'université et l'industrie qui doivent réfléchir aux besoins futurs en termes de compétences, depuis l'apprentissage machine jusqu'à la science des données (UKRS, p9).

Au sujet de la connaissance, la version préliminaire de la Déclaration de Montréal formule le principe suivant : « Le développement de l'IA devrait promouvoir la pensée critique et nous prémunir

contre la propagande et la manipulation ». Si l'éveil de la pensée critique fait dans une certaine mesure écho aux notions de littératie numérique développées à divers degrés et de différentes manières dans les rapports, la Déclaration de Montréal se focalise sur la protection du public face à la propagande et à la manipulation, tandis que les notions d'épanouissement, de liberté, de puissance et de pouvoir surgissent plus fortement dans les autres rapports. Pour plusieurs d'entre eux, la connaissance n'apparaît pas seulement comme un rempart, mais aussi comme une porte ouverte sur de nombreuses possibilités futures.

DÉMOCRATIE

On retrouve la valeur (ou la notion) de démocratie dans tous les rapports et dans des proportions comparables. Les recommandations qui évoquent la démocratie sont notamment associées à : gouvernance, collectivisme/individualisme, gouvernance démocratique, communs numériques, vie privée et confidentialité.

Un premier thème a trait à la gouvernance. Comme on l'a déjà vu avec la valeur d'autonomie, les rapports insistent sur l'idée que l'IA doit rester sous le contrôle humain (AI Now, p.1). D'où la nécessité de créer un cadre de supervision spécialisé (IEEE, p.22 ; UKRS, p.12) ou des systèmes d'audit (CNIL, p.57). Faut-il laisser le secteur privé s'autoréguler ? La réponse qui ressort de la lecture des rapports est plutôt négative – mais il faut reconnaître que le point de vue opposé n'est guère présent, car la seule compagnie dont on peut lire les principes/recommandations (Google) n'évoque pas cette question. En ce qui concerne le type de gouvernance, certaines recommandations laissent transparaître une logique « top down » assez classique : par exemple, chez IEEE ou UKRS, avec l'idée qu'il faut rechercher l'acceptabilité sociale ou « consulter » les citoyens (IEEE, p.31). Asilomar, pour sa part, évoque le dialogue nécessaire entre les chercheurs et les politiques (*policy-makers*). Les conceptions plus radicales ou directes de la démocratie n'apparaissent pas explicitement dans les rapports.

Quoi qu'il en soit, tous conviennent qu'il faut réguler le développement de l'IA – Villani précise même qu'il faut par exemple prévoir un cadre spécial pour protéger les jeux de données les plus sensibles (Villani, p.20). Mais ce qui donne à ces recommandations une dimension proprement démocratique, c'est que l'encadrement ou le contrôle en question se doit d'être transparent. AI Now préconise ainsi que les systèmes d'IA utilisés dans les agences publiques soient disponibles pour des audits, des tests et des révisions publiques (AI Now, p.1). L'idée d'un « corps public d'experts » qui contrôlerait les algorithmes « pour vérifier par exemple qu'ils n'opèrent pas de discrimination » se retrouve également dans CNIL (CNIL, p.58), laquelle va d'ailleurs plus loin qu'AI Now puisque la mission de ces experts ne semblerait pas se cantonner au secteur public. Puisqu'il gêne la transparence, le problème de l'opacité algorithmique est souvent mentionné. Villani remarque ainsi qu'être en mesure « d'ouvrir les boîtes noires » tient de l'enjeu démocratique (Villani, p.21). Il convient donc de soutenir la recherche dans le domaine de l'explicabilité des algorithmes (Villani, p.21).

La démocratie résonne aussi dans les appels à la diversité – culturelle, sociale et de genre, précise la CNIL (CNIL, p.55) – chez les concepteurs d'algorithmes puisqu'il est improbable qu'un sous-groupe (habituellement des hommes blancs riches) puisse anticiper et répondre adéquatement aux besoins de tous les membres de la société. Villani souhaite dès lors une IA « inclusive et diverse » (Villani, p.22) tandis que l'IEEE (IEEE, p.27) recommande aux concepteurs et développeurs d'avoir conscience de la diversité des normes culturelles existantes parmi les utilisateurs des systèmes d'IA. Pour Google, enfin, c'est un des rôles des compagnies que de partager les connaissances et de démocratiser ainsi l'IA afin que plus de personnes développent des applications utiles (Google).

La version préliminaire de la Déclaration de Montréal propose comme principe : « Le développement de l'IA devrait favoriser la participation éclairée à la vie publique, la coopération et le débat démocratique. » On ne s'étonnera pas de l'absence des enjeux de diversité qui sont pris en charge, dans la Déclaration de Montréal, par le principe de justice. On peut

toutefois se demander si les enjeux de gouvernance et de transparence n'auraient pas leur place au sein de ce principe. En particulier, le principe de démocratie de la Déclaration de Montréal reste muet sur la question de savoir qui devrait contrôler le développement de l'IA et comment devrait s'opérer le partage entre la gouvernance publique et privée, experte et populaire.

RESPONSABILITÉ

On trouve des recommandations liées à la responsabilité dans tous les rapports, et les thématiques qui ressortent le plus sont : sécurité et intégrité des systèmes, justice sociale, compétences de l'IA, partage de la responsabilité, imputabilité et responsabilité partagée.

C'est d'abord l'enjeu de la prise de décision qui touche à la notion de responsabilité : lorsqu'une IA peut agir seule, quand doit-elle être surveillée ou complétée par un humain (AI Now, p.1) ? Pour certains, une machine ne doit jamais prendre de décision seule (c'est-à-dire sans intervention humaine) si cela a des conséquences sérieuses pour les personnes (CNIL, p.45).

Afin d'attribuer correctement la responsabilité à l'une ou l'autre entité (ou aux deux), il faut s'assurer que les humains interagissant avec des IA aient les formations nécessaires pour comprendre, avoir un esprit critique, et mesurer les limites et les biais qu'ils vont devoir corriger. Certaines recommandations vont plus loin en suggérant que, dès lors que l'IA est susceptible de reproduire des biais et des discriminations, et à mesure que son irruption dans nos vies sociales et économiques s'accélère, être en mesure « d'ouvrir les boîtes noires » devient une question de démocratie (Villani, p.21). Si les enjeux de compétitivité laissent présager que les entreprises ne pourront pas toutes, ou pas tout le temps, fournir de la transparence absolue, en revanche, il est plusieurs fois recommandé que l'utilisation de systèmes d'IA dans la sphère publique soit la plus transparente possible. D'abord, en ne faisant pas appel à des entreprises privées pour gérer les systèmes publics (AI Now, p.1), ensuite en soumettant les systèmes publics aux plus stricts tests, évaluations, audits, inspections, et standards de responsabilité (AI Now, p.1).

Être responsable, c'est aussi anticiper les problèmes : comment éviter les écueils, quelles infrastructures mettre en place ? À ce sujet, certaines recommandations sont claires : le principe de vigilance devrait être roi (CNIL, p.50) et les concepteurs d'IA devraient toujours avoir en tête la possible imprévisibilité des algorithmes, ainsi que leur caractère évolutif et autonome. Ce principe de vigilance vise à freiner, ou, du moins, à contrebalancer le risque de confiance excessive en l'IA (CNIL, p.50). De nombreuses pistes sont évoquées, comme la création de systèmes d'enregistrement et de traçabilité afin d'être en mesure de remonter à la source d'un algorithme et de déterminer la responsabilité en cas de problème (IEEE, p.27).

Tous les rapports soulignent qu'actuellement le système judiciaire peine à suivre le rythme effréné des développements de l'ère de la donnée et de l'IA, et par conséquent, à offrir des moyens de réguler ces nouvelles technologies. Il s'agit donc de mobiliser des ressources pour le mettre à jour (Asilomar).

Deux éléments clés semblent nécessaires pour encadrer les systèmes d'IA :

1. l'implication de l'appareil judiciaire pour contrôler, corriger, délimiter et aider ;
2. l'implication de scientifiques indépendants dans la conception des appareils de surveillance, d'appel et de label de tous ces systèmes d'IA.

Ces deux groupes devront travailler ensemble pour établir les bonnes pratiques de test de contrôle (Asilomar).

Les entreprises ne devraient pas pour autant se cantonner dans la passivité. Puisqu'elles doivent s'assurer de ne pas amplifier les biais ou de ne pas réaliser d'erreurs (AI Now, p.1), une part importante du travail à effectuer est du côté de la prévention, par exemple en ayant recours à des versions d'essais avant le lancement global d'une application de l'IA (AI Now, p.1). Ces tests préliminaires devraient vérifier non seulement la manière dont les

algorithmes ont été tissés, mais surtout vérifier les données sur lesquelles ils ont été entraînés (AI Now, p.1). Pour cette raison, il est conseillé d'avoir des informations sur la provenance et la gestion de ces données d'entraînement, ainsi que des sauvegardes pour pouvoir les explorer en cas d'anomalie (AI Now, p.1).

La responsabilité dans le domaine judiciaire est un sujet brûlant, et la responsabilité de prendre la décision la plus appropriée et d'éviter les injustices (d'en créer, de les renforcer) est au cœur de nombreuses discussions. Ainsi, Asilomar recommande que tout système autonome impliqué dans des décisions judiciaires puisse être à même de fournir des explications claires quant au cheminement de la décision. L'idée est que ces explications soient analysées par une personne compétente qui a reçu la formation adéquate pour comprendre les rouages de l'algorithme et que les explications soient intelligibles.

La thématique de la responsabilité concerne aussi la médiation entre le public et les fournisseurs de systèmes d'IA, donc l'ouverture et la transparence. Il faut impliquer toute la société au sujet du débat sur la responsabilité humaine (Villani, p.22). L'esprit critique du public sera mis à contribution d'abord dans les cas de médiations (précédemment abordés) – s'il veut se défendre en cas de litige, de désaccord, il faut que les algorithmes soient explicables, et qu'il puisse les comprendre –, et aussi dans le cadre de consultations publiques et citoyennes ou d'audits nationaux ouverts.

De son côté, la version préliminaire de la Déclaration de Montréal propose comme principe : « Les différents acteurs du développement de l'IA devraient assumer leur responsabilité en œuvrant contre les risques de ces innovations technologiques ». En ces termes, la Déclaration de Montréal englobe l'essence des recommandations proposées par les différents rapports, mais reste très générale (ces derniers pouvant distiller des prescriptions plus précises). La Déclaration de Montréal pourrait exposer plus amplement l'enchevêtrement des acteurs prenant part à l'élaboration de ces systèmes et l'éventail des écueils qu'ils se doivent d'éviter.

3. LES RAPPORTS SUR LE DÉVELOPPEMENT DE L'IA : FICHES TECHNIQUES

3.1 LES SEPT RAPPORTS RETENUS

(AI NOW) AI NOW 2017 REPORT

Sous-titre : non
Date de publication : novembre 2017
Pays : É.-U.
Langue : anglais
Organisation ou signataires : AI Now Institute (rapport signé par Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker, Kate Crawford)
Nombre de pages : 37
Résumé : oui (3 pages)
Principes généraux bien identifiés : non
Recommandations bien identifiées : oui (10)
Thèmes principaux : travail et automatisation, biais et inclusion, droits et liberté, éthique et gouvernance.
Notes : un rapport annuel qui cite beaucoup d'études récentes et semble avoir pour vocation de faire le point sur les avancées de la recherche.
Lien : https://ainowinstitute.org/AI_Now_2017_Report.pdf

(CNIL) COMMENT PERMETTRE À L'HOMME DE GARDER LA MAIN – LES ENJEUX ÉTHIQUES DES ALGORITHMES ET DE L'IA

Sous-titre : Les enjeux éthiques des algorithmes et de l'intelligence artificielle. Synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une république numérique
Date de publication : décembre 2017
Pays : 80
Langue : français
Organisation ou signataires : CNIL : Commission nationale informatique et liberté (préface d'Isabelle Falque-Pierrotin, présidente de la CNIL)
Nombre de pages : 80
Résumé : oui (2 pages)
Principes éthiques généraux bien identifiés : oui (vigilance et loyauté)
Recommandations bien identifiées : oui (6)
Thèmes principaux : les enjeux éthiques de l'IA, les applications par secteur (santé, éducation, vie de la cité et politique, culture et média, justice, banque et finance, sécurité et défense, assurance, emploi et RH).
Notes : un des rapports les plus complets concernant les enjeux éthiques de l'IA.
Lien : https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf

(IEEE) ETHICALLY ALIGNED DESIGN. VERSION 2 – FOR PUBLIC DISCUSSION

Sous-titre : A vision for prioritizing human well-being with autonomous and intelligent systems.
Date de publication : Décembre 2017
Pays : international
Langue : anglais
Organisation ou signataires : IEEE (Institute of Electrical and Electronics Engineers) ; signé par des sous-comités de l'IEEE qui regroupent plusieurs centaines de participants internationaux.
Nombre de pages : 266
Résumé : oui (17 pages)
Principes éthiques généraux bien identifiés : oui (5)
Recommandations bien identifiées : oui
Thèmes principaux : enjeux éthiques, juridiques, politiques ; questions spécifiquement liées aux technologies de l'information et de la communication ; sécurité ; *ethics by design* ; contrôle des données.
Notes : chacun des chapitres a été écrit par des comités d'experts.
Lien : https://ethicsinaction.ieee.org/

(ASILOMAR) ASILOMAR AI PRINCIPLES

Sous-titre : non
Date de publication : 2017
Pays : international
Langue : anglais et traductions disponibles en chinois, allemand, japonais, coréen et russe.
Organisation ou signataires : Future of Life Institute, signé par plus de 1200 chercheurs et 2500 non-chercheurs.
Nombre de pages : document en ligne
Résumé : non
Principes éthiques généraux bien identifiés : oui (23)
Recommandations bien identifiées : non
Thèmes principaux : éthique de la recherche, valeurs morales, enjeux à long terme.
Notes : Il ne s'agit pas d'un rapport, mais d'un ensemble de principes qui proviennent de discussions entre experts lors d'une conférence à Asilomar, en Californie. En 1975, une autre conférence à Asilomar a établi des principes en bioéthique.
Lien : https://futureoflife.org/ai-principles/?cn-reloaded=1

(UKRS) AI IN THE UK : READY, WILLING, AND ABLE?

Sous-titre : non
Date de publication : 16 avril 2018
Pays : Royaume-Uni
Langue : anglais
Organisation ou signataires : Parlement (House of Lords) ; comité de 13 personnes.
Nombre de pages : 184
Résumé : oui (5 pages)
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : oui (73)
Thèmes principaux : questions d'éthique et d'économie politique (« innover en IA »). Impact de l'IA sur différents secteurs : économie, travail, éducation, santé, justice.
Notes : le rapport est divisé en 420 paragraphes dont l'auteur est souvent identifié en note.
Lien : https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

(VILLANI) DONNER UN SENS À L'INTELLIGENCE ARTIFICIELLE

Sous-titre : Pour une stratégie nationale et européenne
Date de publication : 8 mars 2018
Pays : France
Langue : français
Organisation ou signataires : Missions parlementaires confiées au député Cédric Villani et à 6 autres parlementaires.
Nombre de pages : 235
Résumé : oui (15 pages)
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : non
Thèmes principaux : questions d'éthique et d'économie politique, politique de la recherche, impact sur l'emploi et dans les secteurs de l'éducation, la santé, l'agriculture, le transport et la défense.
Lien : http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/184000159.pdf

(GOOGLE) AI AT GOOGLE : OUR PRINCIPLES

Sous-titre : non
Date de publication : 7 juin 2018
Pays : É.-U.
Langue : anglais
Organisation ou signataires : Google, présenté par son CEO Sundar Pichai
Nombre de pages : document en ligne
Résumé : non
Principes éthiques généraux bien identifiés : oui (7)
Recommandations bien identifiées : oui (4)
Thèmes principaux : éthique de l'IA
Notes : l'entreprise s'engage à ne pas déployer d'IA dans certains domaines (armement) ou certaines circonstances (à l'encontre des droits humains).
Lien : https://www.blog.google/technology/ai/ai-principles/

3.2
RAPPORTS EXAMINÉS, MAIS NON RETENUS

A NEXT GENERATION ARTIFICIAL INTELLIGENCE DEVELOPMENT PLAN

Sous-titre : non
Date de publication : Juillet 2017
Pays : Chine
Langue : anglais (traduction)
Organisation ou signataires : State council of the People's Republic of China
Nombre de pages : 28
Résumé : non
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : oui
Thèmes principaux : stratégie nationale pour le développement économique
Lien : https://chinacopyrightandmedia.wordpress.com/2017/07/20/a-next-generation-artificial-intelligence-development-plan/

STRATEGY FOR DENMARK'S DIGITAL GROWTH

Sous-titre : non
Date de publication : 2018
Pays : Danemark
Langue : anglais
Organisation ou signataires : Ministry of Industry, Business and Financial Affairs
Nombre de pages : 68
Résumé : oui (6 pages)
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : oui
Thèmes principaux : stratégie nationale pour le développement économique
Lien : https://em.dk/english/news/2018/01-30-new-strategy-to-make-denmark-the-new-digital-frontrunner

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS

Sous-titre : Artificial Intelligence for Europe
Date de publication : 25 avril 2018
Pays : Union européenne
Langue : anglais
Organisation ou signataires : European Commission
Nombre de pages : 20
Résumé : non
Principes éthiques généraux bien identifiés : oui
Recommandations bien identifiées : oui
Thèmes principaux : stratégie nationale pour le développement économique
Lien : https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

FINLAND'S AGE OF ARTIFICIAL INTELLIGENCE

Sous-titre : Turning Finland into a leading country in the application of artificial intelligence: Objective and recommendations for measures
Date de publication : 18 décembre 2017
Pays : Finlande
Langue : anglais
Organisation ou signataires : Ministry of Economic Affairs and Employment
Nombre de pages : 76
Résumé : oui (3 pages)
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : oui (8)
Thèmes principaux : stratégie nationale pour le développement économique
Lien : http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y

ETHICS COMMISSION AUTOMATED AND CONNECTED DRIVING

Sous-titre : non
Date de publication : Juin 2017
Pays : Allemagne
Langue : anglais
Organisation ou signataires : Federal Ministry of Transport and Digital Infrastructure
Nombre de pages : 36
Résumé : non
Principes éthiques généraux bien identifiés : oui
Recommandations bien identifiées : oui
Thèmes principaux : Éthique des véhicules autonomes
Lien : https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

NATIONAL STRATEGY FOR ARTIFICIAL INTELLIGENCE #AIFORALL

Sous-titre : Discussion paper
Date de publication : Juin 2018
Pays : Inde
Langue : anglais
Organisation ou signataires : NITI Aayog
Nombre de pages : 115
Résumé : oui (3)
Principes éthiques généraux bien identifiés : oui
Recommandations bien identifiées : oui
Thèmes principaux : stratégie nationale pour le développement économique et sociétal
Lien : http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

ARTIFICIAL INTELLIGENCE AT THE SERVICE OF CITIZENS

Sous-titre : non
Date de publication : Mars 2018
Pays : Italie
Langue : anglais
Organisation ou signataires : The Agency for Digital Italy
Nombre de pages : 79
Résumé : oui (5 pages)
Principes éthiques généraux bien identifiés : oui
Recommandations bien identifiées : oui
Thèmes principaux : impact de l'IA sur la société et dans l'administration publique pour promouvoir le changement

ARTIFICIAL INTELLIGENCE TECHNOLOGY STRATEGY

Sous-titre : Report of Strategic Council for AI Technology
Date de publication : 31 mars 2017
Pays : Japon
Langue : anglais
Organisation ou signataires : Strategic Council for AI Technology
Nombre de pages : 25
Résumé : non
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : oui
Thèmes principaux : stratégie nationale pour le développement de l'IA
Lien : http://www.nedo.go.jp/content/100865202.pdf

TOWARDS AN AI STRATEGY IN MEXICO

Sous-titre : Harnessing the AI Revolution
Date de publication : Juin 2018
Pays : Mexique
Langue : anglais
Organisation ou signataires : British Embassy in Mexico through the Prosperity Fund, Oxford Insights, C Minds
Nombre de pages : 52
Résumé : oui (3 pages)
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : oui (21)
Thèmes principaux : stratégie nationale pour le développement économique
Lien : https://docs.wixstatic.com/ugd/7be025_e726c582191c49d2b8b6517a590151f6.pdf

SHAPING A FUTURE NEW ZEALAND

Sous-titre : An Analysis of the Potential Impact and Opportunity of Artificial Intelligence on New Zealand's Society and Economy
Date de publication : Mai 2018
Pays : Nouvelle-Zélande
Langue : anglais
Organisation ou signataires : AI Forum of New Zealand
Nombre de pages : 108
Résumé : oui (5 pages)
Principes éthiques généraux bien identifiés : oui
Recommandations bien identifiées : oui (14)
Thèmes principaux : stratégie nationale pour le développement économique

ARTIFICIAL INTELLIGENCE IN SWEDISH BUSINESS AND SOCIETY

Sous-titre : Analysis of development and potential
Date de publication : Mai 2018
Pays : Suède
Langue : anglais
Organisation ou signataires : Vinnova
Nombre de pages : 32
Résumé : non
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : oui
Thèmes principaux : développement économique et services publics
Lien : https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr_18_09.pdf

INDUSTRIAL STRATEGY

Sous-titre : AI Sector Deal
Date de publication : Avril 2018
Pays : R.-U.
Langue : anglais
Organisation ou signataires : Gouvernement
Nombre de pages : 21
Résumé : oui (3 pages)
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : oui
Thèmes principaux : stratégie nationale pour le développement économique
Lien : https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal__4_.pdf

PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE

Sous-titre : non
Date de publication : Octobre 2016
Pays : États-Unis
Langue : anglais
Organisation ou signataires : Executive Office of the President, National Science and Technology Council Committee on Technology
Nombre de pages : 58
Résumé : oui (4)
Principes éthiques généraux bien identifiés : oui
Recommandations bien identifiées : oui (23)
Thèmes principaux : état actuel de l'IA, applications présentes et futures, questions soulevées pour la société
Lien : https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN

Sous-titre : non
Date de publication : Octobre 2016
Pays : États-Unis
Langue : anglais
Organisation ou signataires : National Science and Technology Council, Networking and Information Technology Research and Development Subcommittee
Nombre de pages : 48
Résumé : oui (2 pages)
Principes éthiques généraux bien identifiés : non (quelques-uns)
Recommandations bien identifiées : oui (7)
Thèmes principaux : objectifs pour la recherche en IA financée par le gouvernement fédéral
Lien : https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf

ARTIFICIAL INTELLIGENCE, AUTOMATION, AND THE ECONOMY

Sous-titre : non
Date de publication : Décembre 2016
Pays : États-Unis
Langue : anglais
Organisation ou signataires : Executive Office of the President
Nombre de pages : 55
Résumé : oui (4 pages)
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : oui (3)
Thèmes principaux : impacts de l'automation par l'IA sur l'économie et stratégies pour accroitre les bénéfices
Lien : https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

SUMMARY OF THE 2018 WHITE HOUSE SUMMIT ON ARTIFICIAL INTELLIGENCE FOR AMERICAN INDUSTRY

Sous-titre : non
Date de publication : 10 mai 2018
Pays : États-Unis
Langue : anglais
Organisation ou signataires : The White House Office of Science and Technology Policy
Nombre de pages : 15
Résumé : oui (1 page)
Principes éthiques généraux bien identifiés : non
Recommandations bien identifiées : oui
Thèmes principaux : stratégie nationale pour le développement économique
Lien : https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf

3.3 AUTRES RAPPORTS CONSULTÉS

(SUÈDE) NATIONAL APPROACH FOR ARTIFICIAL INTELLIGENCE

https://www.regeringen.se/49a828/contentassets/844d30fb0d594d1b9d96e2f5d57ed14b/2018ai_webb.pdf

(ALLEMAGNE) ECKPUNKTE DER BUNDESREGIERUNG FÜR EINE STRATEGIE KÜNSTLICHE INTELLIGENZ

https://www.bmwi.de/Redaktion/DE/Downloads/E/eckpunktepapier-ki.pdf?__blob=publicationFile&v=4

(FINLANDE) WORK IN THE AGE OF ARTIFICIAL INTELLIGENCE

http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160931/19_18_TEM_Tekoalyajan_tyo_WEB.pdf

(CHINE) THREE-YEAR ACTION PLAN TO PROMOTE THE DEVELOPMENT OF NEW-GENERATION ARTIFICIAL INTELLIGENCE INDUSTRY

http://www.miit.gov.cn/n1146295/n1652858/n1652930/n3757016/c5960820/content.html

(AUSTRALIE) AUSTRALIA 2030: PROSPERITY THROUGH INNOVATION

https://www.industry.gov.au/sites/g/files/net3906/f/May%202018/document/pdf/australia-2030-prosperity-through-innovation-full-report.pdf

Merci à Paloma Fernandez-McAuley pour son aide.

< >
Déclaration de Montréal
IA responsable_
</ >

PARTIE 3

RAPPORT DES RÉSULTATS DES ATELIERS DE COCONSTRUCTION DE L'HIVER

TABLE DES MATIÈRES

TABLE DES FIGURES ET DES TABLEAUX

RÉDACTION

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

CAMILLE VÉZY, candidate au doctorat en communication, Université de Montréal

VALENTINE CROSSET, doctorante en criminologie, Université de Montréal

ALESSIA ZARZANI, Ph.D en aménagement, Université de Montréal et Ph.D en Paysage et Environnement, Université la Sapienza de Roma

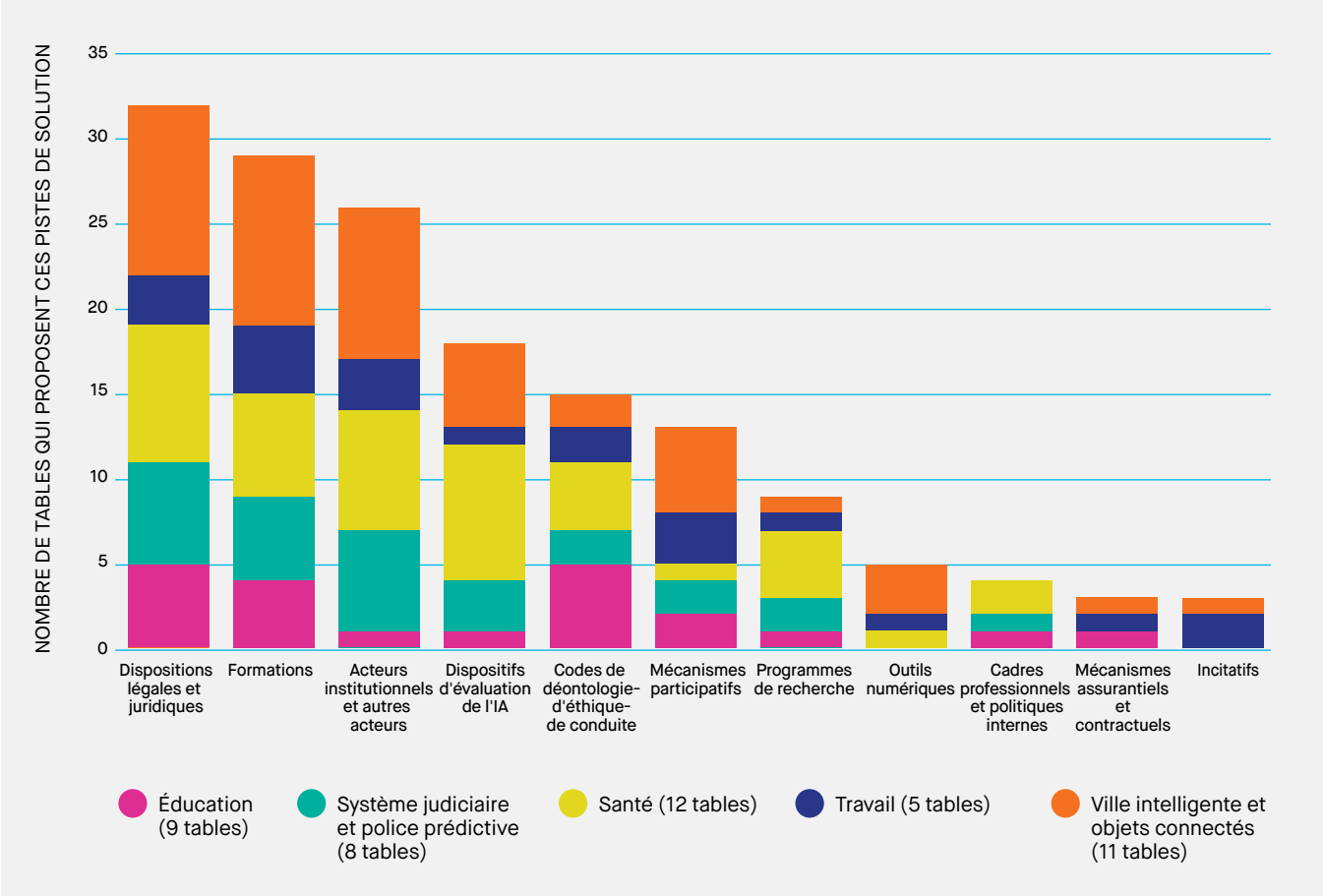
Dans ce document, l'utilisation du genre masculin a été adoptée afin de faciliter la lecture et n'a aucune intention discriminatoire.

1. RÉSUMÉ

Les citoyens se sont réunis autour de 45 tables pour discuter de leur perception des enjeux et des risques liés au développement responsable de l'intelligence artificielle (IA).

Ils ont formulé 11 catégories de pistes de solution pour répondre à ces enjeux ou encadrer ces risques :

Tableau 1 : Les pistes de solution proposées pour répondre aux enjeux identifiés



2. LES DONNÉES DE LA COCONSTRUCTION : NOTES EXPLICATIVES

La présente section fait état des résultats collectés lors des tables de coconstruction qui ont eu lieu à l'hiver 2018 dans le cadre de la ***Déclaration de Montréal pour un développement responsable de l'intelligence artificielle,*** soit 45 tables qui ont réuni plusieurs centaines de citoyens. Les discussions se sont organisées autour de 5 grands secteurs de développement de l'IA : le secteur de l'éducation (9 tables) ; le secteur judiciaire et police prédictive (8 tables) ; le secteur de la santé (12 tables), le secteur du travail (5 tables), et le secteur ville intelligente et objets connectés (11 tables). L'analyse présentée dans cette section a également été enrichie par les discussions de différentes activités satellites (intervention lors de cours ; cafés-citoyens portant sur les mêmes thématiques mais ne suivant pas strictement la méthode utilisée lors des tables de coconstruction).

Pour la compréhension de cette section, il est à noter que les discussions ont porté sur les enjeux du développement responsable de l'IA, mais aussi sur ceux liés à la gestion des données (en particulier, les données personnelles et les données massives) — qu'il s'agisse de données sur lesquelles les algorithmes apprennent, ou de celles qui font, d'une manière ou d'une autre, l'objet d'une analyse par une IA. Ces enjeux étant étroitement liés, ils ont été considérés conjointement pour la présente analyse.

Des scénarios déclencheurs ont servi de base aux discussions afin de collecter deux types de données : la perception des citoyens des enjeux et des risques du développement de l'IA et leurs pistes de solution pour y répondre (cf. scénarios partie 1, Section 6, Annexe 2).

Pour les fins de cette section, l'analyse est restée descriptive et auplus proche de la parole des citoyens. Les grandesdirections attendues (ou attentes normatives) entermes de développement responsable de l'IA réfèrent aux recommandations des citoyens qui ne se sont pas précisées en pistes de solution, icientendues comme mécanismes concrets. Ellespermettent néanmoins de dégager les principalespositions des citoyens face au développement de l'IA. Lorsque ces attentes normatives ont été misesen tension lors des discussions ou lorsque lescitoyens ont considéré que répondre à ces attentesétait un enjeu, elles ont été considérées dans lacatégorie des enjeux.

Chaque table de coconstruction a été invitée à choisir 2 ou 3 enjeux à traiter en priorité d'ici 2025. Seuls les enjeux que les citoyens ont considérés comme prioritaires ont été pris en compte pour la section 3.1. Ces enjeux prioritaires ont été décrits sur la base des formulations citoyennes et classés,

pour chaque secteur, en fonction des principes de la première version de la Déclaration auxquels ils sont reliés. Cependant, il est à noter que ce n'est pas parce que des enjeux n'ont pas été considérés comme prioritaires qu'ils n'ont pas été abordés, qu'ils sont moins importants ou que tous les principes n'ont pas été abordés pour chacun des secteurs. Un seul principe par secteur (celui qui ressort comme étant le plus essentiel) est détaillé dans cette section.

L'ensemble des discussions a fait l'objet d'une analyse thématique réalisée par le biais du logiciel NVivo. L'analyse thématique a été réalisée afin de mettre en évidence l'étendue des risques et enjeux du développement responsable de l'IA selon la perception des citoyens (cf. cartographie des enjeux p.117).
Ces enjeux sont regroupés en 12 catégories et ne sont pas mutuellement exclusives. Nous reconnaissons qu'il ne s'agit que d'une façon parmi d'autres de classer les différentes discussions qui ont eu lieu. Quant aux pistes de solution identifiées par les citoyens pour répondre à ces enjeux, elles ont été classées en 11 principales catégories. Ces catégories sont mutuellement exclusives ; ce qui a permis de les accompagner de données quantitatives.

Concernant les données quantitatives de ce rapport, le nombre d'occurrences correspond au nombre de tables où chaque enjeu ou piste de solution a été formulé de façon consensuelle, conformément au processus de coconstruction. Le nombre total de pistes de solution (n=190) correspond à celles identifiées comme prioritaires par les citoyens (car ils ont été invités à les formuler clairement sur des affiches). Cependant, d'autres pistes de solutions ont été discutées et prises en considération lors de l'écriture des différentes sections subséquentes. S'il n'a pas été possible de comptabiliser ces dernières, elles ont permis d'enrichir le présent rapport et de préciser certains mécanismes.

Les citations du rapport sont présentées de façon à faire référence à la table de coconstruction lorsqu'elles sont issues de formulations de groupe (consensus). Les autres citations correspondent à des formulations individuelles (rédigées sur des post-it par les participants ou retranscrites en verbatim par les membres de l'équipe).

3. LES GRANDES DIRECTIONS ATTENDUES PAR LES CITOYENS

De manière générale, les participants ont reconnu que l'avènement de l'IA s'accompagne d'importants bénéfices potentiels. Notamment, dans le secteur judiciaire ou celui du travail, les participants ont reconnu le gain de temps que pourrait offrir le recours à des dispositifs d'IA :

« Cela permettrait de réduire les temps d'attente du traitement des dossiers. »

Cependant, les citoyens ont mentionné que le développement de l'IA doit se faire avec prudence, et dès à présent, afin de prévenir les dérives, bien que certains considèrent les possibilités qu'amène l'IA comme encore limitées. La mise en place d'un encadrement est ainsi reconnue comme nécessaire afin de prévenir les risques plutôt que de déterminer qui blâmer lorsqu'ils se produisent :

« Tu ne veux pas tant savoir qui poursuivre quand ça va mal, tu veux plutôt trouver des moyens contraignants pour t'assurer que cela n'aille pas mal. »

Les participants ont ainsi mis en avant la nécessité de mettre en place différents mécanismes pour assurer la qualité, l'intelligibilité, la transparence et la pertinence des informations transmises. Ils ont également souligné la difficulté à garantir un véritable consentement éclairé.

La grande majorité des participants a reconnu la nécessité de faire cadrer les intérêts publics avec les intérêts privés et d'empêcher l'apparition de monopole, voire de limiter l'influence des entreprises (parfois considérées comme ingouvernables) par des mesures plus coercitives et légalistes que pour les autres enjeux identifiés. Ces mécanismes devraient être, dans la mesure du possible, simples et évolutifs afin de pouvoir s'adapter au rythme du développement de l'IA et permettre son contrôle régulier. Dans le secteur judiciaire, certains participants ont parlé d'un « gouffre » séparant la technologie (définie comme rapide, innovante, voire abstraite) et nos institutions (souvent trop rigides dans leur intégration technologique) qui n'arrivent pas à faire face à ces changements de société. Certaines tables sont allées jusqu'à proposer une « nationalisation de l'IA » qui deviendrait alors « un service public et les informaticiens, des fonctionnaires. » (Table ville intelligente et objets connectés, INM, Montréal, 18 février 2018, scénario Réfrigérateur connecté).

Les participants recommandent également de garantir une approche contextuelle de l'IA, qui doit tenir compte de différents paramètres (p. ex., collecte obligatoire ou facultative des données sur lesquelles apprennent les algorithmes). Ces mécanismes devraient émaner et impliquer des personnes formées et indépendantes, favoriser la diversité et l'intégration des plus vulnérables, et protéger la mixité des modes de vie.

Quelles que soient les applications, la majorité des participants souligne le fait que l'IA doit rester un outil et que la décision finale doit rester celle d'un humain (qu'il s'agisse d'une décision de justice, d'une décision concernant l'embauche ou d'un diagnostic en santé), ce qui implique de reconnaître ses limites.

« L'IA propose, l'humain dispose. »

La protection de la vie privée des individus et la gestion des données personnelles ont été largement discutées. Par exemple, le traitement des données de santé devrait faire l'objet d'une gestion toute particulière vu le caractère sensible des informations. Elle devrait ainsi favoriser à la fois des méthodes de contrôle hiérarchisé selon le type d'usage et adopter la sécurité comme mode opérationnel. Concernant le secteur du monde du travail, les participants ont recommandé l'obligation d'informer les usagers du traitement de leurs données.

Conscients que ces recommandations impliquent des changements institutionnels importants, des participants ont souligné la nécessité de garder en tête que l'IA n'est pas forcément souhaitable a priori.

« Just because you can doesn't mean you should. »

Ainsi, les citoyens se sont généralement accordés pour dire que les conséquences des usages de l'IA dans les différents secteurs — pour l'individu comme pour la société dans son ensemble — doivent clairement être mesurées afin de mettre en place des balises sans pour autant limiter indûment le progrès.

4. LA PERCEPTION CITOYENNE DES ENJEUX DU DÉVELOPPEMENT RESPONSABLE DE L'IA

4.1. INTRODUCTION

Les citoyens ayant participé aux journées de coconstruction ont été invités à sélectionner 2 ou 3 enjeux à adresser en priorité d'ici 2025 concernant le développement responsable de l'IA et à les mettre en lien avec les principes de la première version de la Déclaration.

Tableau 2 : Enjeux prioritaires identifiés par les citoyens en fonction des principes de la Déclaration (nombre de tables).

	Éducation	Système judiciaire et police prédictive	Monde du travail	Santé	Ville intelligente et objets connectés	Nombre total de tables qui considèrent ces enjeux comme prioritaires
Responsabilité	6	5	3	10	5	29
Autonomie	7	3	2	5	9	26
Vie privée	6	5	1	9	4	25
Bien-être	6	4	2	6	5	23
Connaissance	6	5	4	4	2	21
Justice	6	4	5	4	4	21
Démocratie	1	4	3	1	7	16
Nombre total de tables de coconstruction	9	8	5	12	11	45

Le principe de responsabilité a été celui qui a été jugé le plus souvent prioritaire, suivi du principe d'autonomie, de celui de vie privée puis de ceux de bien-être (individuel et collectif), de connaissance et de justice. Il faut cependant noter qu'ils sont tous étroitement liés.

Les principes de connaissance, de responsabilité, de vie privée, de justice et de démocratie sont présentés ci-dessous par secteurs. Pour ce qui est du principe d'autonomie, très souvent choisi comme prioritaire, il a trait à la préservation, voire l'encouragement de l'autonomie individuelle face

à des risques de déterminisme technologique et de dépendance aux outils. Il soulève également l'enjeu d'une double liberté de choix : pouvoir suivre son propre choix face à une décision orientée par l'IA, mais également pouvoir choisir de ne pas utiliser ces outils sans pour autant risquer une exclusion sociale. La liberté comprise dans ce principe d'autonomie par rapport à des systèmes d'IA relèverait ainsi d'une capacité d'autodétermination de toute personne.

« Développer des technologies qui favorisent l'autonomie humaine et la liberté de choix. »

(Table éducation, bibliothèque de Laval, 24 mars 2018, scénario Hyperpersonnalisation de l'éducation).

Le principe de bien-être occupe également une place importante pour les participants. Il est présent en filigrane à toutes les tables, manifestant un souhait collectif d'avancer vers une société juste, équitable et favorisant le développement de tous. Le bien-être est ainsi un enjeu à la fois collectif (lié aux enjeux d'accessibilité et d'équité présents dans le principe de justice) et individuel, visant l'épanouissement de chacun sans entrave à l'autonomie et la vie privée. Les participants ont ainsi manifesté une préférence pour un développement de l'IA « qui permette l'épanouissement personnel et social de tout individu. » (Table éducation, bibliothèque Père Ambroise, Montréal, 3 mars 2018, scénario AlterEgo).

De façon générale, le principe de bien-être a également pris la forme d'un appel au maintien d'une relation humaine et émotionnelle de qualité entre experts et usagers dans tous les secteurs.

LES PRINCIPAUX ENJEUX ABORDÉS PAR SECTEUR

ÉDUCATION

En ce qui concerne le secteur de l'éducation, les enjeux relatifs aux principes de vie privée, de responsabilité, de bien-être et de connaissance ont été considérés comme prioritaires par 6 tables sur 9. Les discussions portant sur les enjeux relatifs au principe de connaissance ont été particulièrement pertinentes pour aborder les questions de transformation des habiletés humaines à l'heure de l'IA :

ENJEUX RELATIFS AU PRINCIPE DE CONNAISSANCE
(6 tables sur 9)

Les enjeux relatifs au principe de Connaissance pour le thème de l'éducation relèvent d'enjeux de transformation des compétences dans un contexte où changent à la fois le métier enseignant et les manières de développer des connaissances et d'y accéder. Ce principe a ainsi surtout été discuté en rapport avec la transformation de la relation d'apprentissage, relevant alors d'un enjeu d'expertise de l'enseignant dont le travail sera amené à être modifié. Il a également été mentionné en lien avec un principe de diversité pour évoquer la nécessité d'entretenir une diversité des intelligences et des rapports au savoir.

« Redéfinition/transformation de la nature de la relation entre l'enseignant et les étudiants dans l'espace pédagogique et modification des rapports au savoir. »

(Table de la SAT, Montréal, 13 mars 2018, scénario Nao).

« Compétences/habiletés humaines : importance de développer plusieurs environnements d'apprentissage. »

(Table du Musée de la civilisation, Québec, 6 avril 2018, scénario AlterEgo).

SYSTÈME JUDICIAIRE ET POLICE PRÉDICTIVE

En ce qui concerne le secteur judiciaire et de la police prédictive, les enjeux relatifs aux principes de vie privée, de responsabilité et de connaissance ont été considérés comme prioritaires par 5 tables sur 8. Les discussions portant sur les enjeux relatifs au principe de responsabilité permettent de préciser l'étendue du principe :

ENJEUX RELATIFS AU PRINCIPE DE RESPONSABILITÉ
(5 tables sur 8)

Le principe de responsabilité s'est formulé de deux principales façons. D'abord au nom d'une revendication pour asseoir la responsabilité humaine en matière de décision judiciaire, et par souci d'imputabilité de la décision (et de toute erreur potentielle). Puis, par le manque de transparence des algorithmes qui vient, pour les citoyens, défier l'imputabilité puisqu'il est difficile de retracer ce qui est pris en compte dans la décision. Le principe de responsabilité est ainsi lié aux principes de connaissance et transparence en ce qui concerne la revendication de rendre explicables les décisions et de préserver les compétences et la place des acteurs humains dans le système judiciaire.

« [La justice] doit rester un outil dans le seul but de protéger les individus. Promotion d'une justice empathique et équitable prenant en compte les singularités et les expériences. L'intelligence artificielle ne doit pas avoir le droit de porter un jugement sur un comportement humain. La décision finale doit toujours comporter une intervention humaine. »

(Table de la SAT, Montréal, 13 mars 2018, scénario Arrestation préventive).

« Transparence, imputabilité et responsabilité quant à la création de l'outil, aux données utilisées et aux conséquences de l'outil. »

(Table de la SAT, Montréal, 13 mars 2018, scénario Libération conditionnelle).

Concernant la responsabilité, c'est parfois la non-prise en compte de l'humain et de son « agentivité » qui a été soulevée. La non-prise en compte de la dynamique humaine et de ses possibilités de changement démontre une inquiétude face à la vision « statique » de l'être humain donnée par l'algorithme, qui rendrait ses décisions problématiques et peu fiables. Dans cet atelier, les participants étaient prêts à faire de « l'agentivité » un principe de la Déclaration.

« Il faut prendre en compte le dynamisme individuel. La possibilité de chacun de pouvoir changer, de pouvoir modifier sa propre trajectoire. »

SANTÉ

En ce qui concerne le secteur de la santé, les enjeux relatifs aux principes de vie privée et de responsabilité ont été considérés comme prioritaires, respectivement par 9 et 10 tables sur 12. Les enjeux relatifs au principe de vie privée revêtent une importance particulière dans ce secteur considérant le côté relativement sensible et le caractère presque toujours personnel des données de santé.

ENJEUX RELATIFS AU PRINCIPE DE VIE PRIVÉE
(9 tables sur 10)

Les participants ont identifié différents enjeux touchant l'atteinte à la vie privée et à la confidentialité. Ces enjeux concernent une possible invasion dans la vie privée qui peut être liée au développement et à la configuration des systèmes d'IA (qui devrait permettre d'éviter le piratage, les pannes et les abus). Ils ont aussi trait à ce que les citoyens ont appelé « la rétroaction » (utilisation de données collectées antérieurement dans un autre but) et l'accès à ces données par des compagnies privées. Face à ces enjeux, les citoyens se sont inquiétés de la façon de s'assurer que les données ne soient pas marchandées et garantir que le patient garde le contrôle sur ses données (en particulier lorsqu'il s'agit de données personnelles), voire qu'il en détienne impérativement la propriété.

« Jusqu'où sommes-nous prêts à partager nos données (informations) personnelles à titre d'individus dans l'optique d'en nourrir des services de santé ? »

(Table du Musée de la civilisation, Québec, 6 avril 2018, scénario Jumeaux numériques).

MONDE DU TRAVAIL

En ce qui concerne le secteur du monde du travail, les enjeux relatifs au principe de justice et de connaissance ont été considérés comme prioritaires (respectivement 5 et 4 tables sur 5). Toutes les tables qui se sont réunies autour du thème du développement de l'IA dans le monde du travail ont ainsi considéré que les enjeux relatifs à la justice, à l'équité ou la diversité devaient être abordés expressément.

ENJEUX RELATIFS AU PRINCIPE DE JUSTICE
(5 tables sur 5)

Le principe de justice fait l'objet de deux préoccupations principales : assurer un partage équitable des bénéfices de l'IA entre tous les acteurs, groupes sociaux et territoires, et « mettre en place des algorithmes non discriminatoires qui favorisent la diversité, l'inclusion et la justice sociale. » (Table du Musée de la civilisation, Québec, 6 avril 2018, scénario L'IA comme passage obligé vers l'emploi).

« Partage des bénéfices de l'IA (gains de productivité); équité entre les groupes sociaux, territoires (villes et régions), prise en compte des vulnérabilités; sens du travail dans la société et dans la construction de nos identités. »

(Table du Musée de la civilisation, Québec, 6 avril 2018, scénario Une restructuration socialement responsable).

VILLE INTELLIGENTE ET OBJETS CONNECTÉS

En ce qui concerne le secteur de la ville intelligente et des objets connectés, les enjeux relatifs aux principes d'autonomie et de démocratie ont été considérés comme prioritaires par 9 et 7 tables sur 11. De nombreux enjeux semblent pouvoir porter atteinte au principe de démocratie selon les citoyens :

ENJEUX RELATIFS AU PRINCIPE DE DÉMOCRATIE
(7 tables sur 11)

Les participants ont discuté d'enjeux liés à l'équilibre entre intérêts collectifs et besoins individuels ; à la gestion de l'accès à l'espace public et au partage de cet espace ou, encore, au partage des bénéfices issus du développement de technologies d'IA (notamment, entre particuliers, secteur public et secteur privé). Ils ont souligné la nécessité et la difficulté d'assurer la prise de décision collective (incluant les citoyens) et éclairée (ce qui implique une certaine transparence concernant le développement des systèmes d'IA) pour définir les lignes directrices visant à mettre en place ou à régler des objets connectés. Les citoyens ont également remis en question la réelle indépendance des pouvoirs publics face au développement de l'IA et ont mis en avant le risque de normalisation de comportements qui pourraient conduire à une marginalisation, risquant ainsi de porter atteinte au principe de démocratie.

« Comment peut-on gérer de façon démocratique un système de transport intelligent ? »

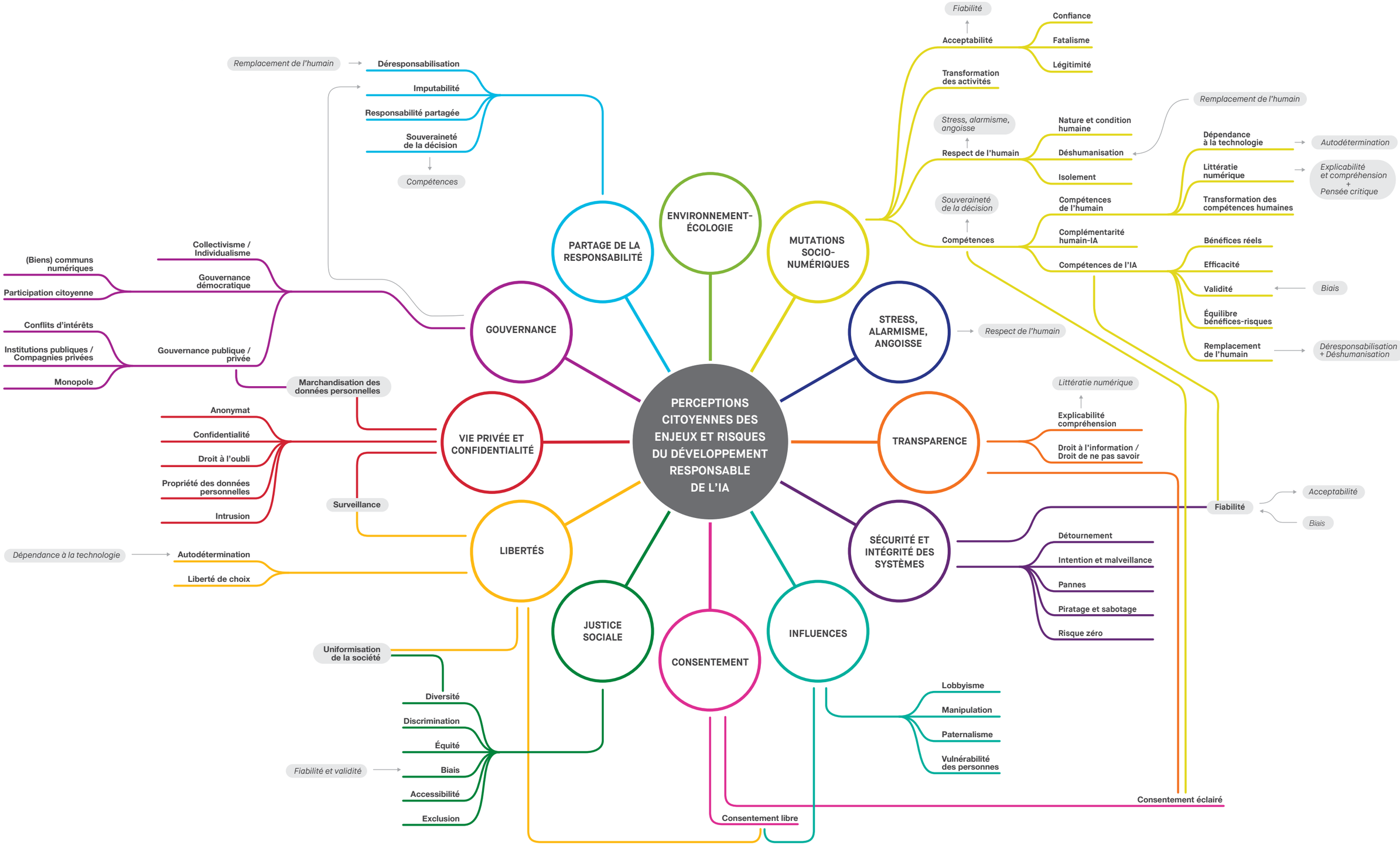
(Table de la bibliothèque Du Boisé, Montréal, 17 mars 2018, scénario Voiture autonome).

4.2.

LES GRANDES CATÉGORIES DE RISQUES ET ENJEUX DU DÉVELOPPEMENT RESPONSABLE DE L'IA

Face aux différents scénarios, 12 grandes catégories de risques et enjeux se sont dégagées des discussions citoyennes sur le développement responsable de l'IA. Ces catégories ne sont pas mutuellement exclusives, elles offrent un portrait de la diversité des thèmes qui ont pu être soulevés par les citoyens en ce qui concerne le développement responsable de l'IA et qui méritent une attention particulière dans l'optique de la création de politiques publiques. La carte heuristique ci-après présente ainsi l'étendue et la diversité des enjeux abordés, classés en catégories et sous-catégories. Parfois, des dilemmes ou oppositions marqués sont ressortis des discussions. La section suivante détaille la définition de chacune des catégories, illustrée par des exemples issus directement de la parole citoyenne.

Tableau 3 : Cartographie des enjeux



GOUVERNANCE

COLLECTIVISME VERSUS INDIVIDUALISME

Cette catégorie réfère à un dilemme qui oppose la protection des intérêts, choix ou responsabilités individuels versus celle des intérêts, choix ou responsabilités collectifs. La réponse à ce dilemme constitue un enjeu important qui dépend très fortement d'une position normative sur laquelle aucun consensus n'a été observé.

« Faire en sorte que la technologie de l'IA soit un outil éducatif au service des visées sociales et démocratiques de l'école en tant que bien public. » (Table éducation, SAT, Montréal, 13 mars 2018, scénario Nao).

« Jumeaux numériques : c'est une façon de procéder très libertarienne et qui cause encore une fois une tension entre le bien-être individuel et collectif. »

« On est dans un stade de démocratie où l'individu est mis en avant à un tel niveau que cela va conduire à une dictature. »

« Comment assurer que les voitures autonomes maximisent le bien-être ? Partage de l'espace public ? Comment concilier la sécurité du plus grand nombre avec la sécurité individuelle ? »

« Les intérêts communs peuvent-ils s'accorder avec des intérêts personnels privés en restant éthiques ? »

GOUVERNANCE PUBLIQUE VERSUS PRIVÉE

Les enjeux relatifs au partage de la gestion du développement de l'IA entre institutions publiques et privées, et les risques qui accompagnent ce partage ont également été abordés. Ces enjeux sont souvent abordés sous la forme de questionnement : Quel serait le meilleur partage ? Laquelle des deux gouvernances est la plus appropriée ?

« Qui pilote tout ça ? Quels pouvoirs a l'organisation ou la compagnie sur cet outil ? Sera-t-on dépendants de cette compagnie-là ? Si ça devient une priorité nationale, quels choix vont être fait pour les programmes pédagogiques à partir de l'implantation ? Est-ce que c'est public ? Privé ? C'est tout l'écosystème de l'éducation qui est redessiné. »

Plus précisément, les risques de **conflits d'intérêts, de marchandisation des données personnelles ou de l'apparition d'un monopole ont été soulevés.** Les participants ont en effet souligné le risque de conflit entre les intérêts privés (essentiellement financiers) et les autres intérêts, qui pourrait nuire à l'indépendance de certains acteurs ou de certaines institutions publiques. Le risque de marchandisation des données personnelles réfère aux enjeux associés à la valeur marchande des données, aux limites de la collecte et des profits faits sur celles-ci, particulièrement au regard de la protection de la vie privée. L'apparition d'un monopole privé dans la gouvernance du développement de l'IA est également un sujet d'inquiétude.

« Éviter l'utilisation commerciale ou d'intérêts autres qu'éducatifs en ce qui concerne les données collectées et analysées par AlterEgo »

(Table éducation, bibliothèque Père Ambroise, Montréal, 3 mars 2018, scénario AlterEgo).

« Comment éviter une marchandisation trop poussée des données et des personnes à leur insu ? »

(Table ville intelligente et objets connectés, SAT, Montréal, 13 mars 2018, scénario Jouet intelligent).

« Trop de concentration de pouvoir (GAFAM) ne permettant pas :

- Partage équitable des bénéfices de l'IA

- L'entrée de nouveaux joueurs (nouveaux modèles d'affaires, ex : coop) »

(Table monde du travail, SAT, Montréal, 13 mars 2018, scénario Une restructuration responsable).

GOUVERNANCE DÉMOCRATIQUE

Si la discussion autour de la gouvernance oppose souvent les institutions publiques aux compagnies privées, les enjeux relatifs à une autre alternative ont été soulevés : celle d'une gouvernance participative qui donne directement la main aux citoyens. Par exemple, ont été discutés les enjeux relatifs à la gestion partagée et collective de biens numériques caractérisés par leur libre accès (biens communs numériques) ou ceux relatifs à la place de la participation citoyenne dans la gouvernance actuelle et à venir (qu'il s'agisse de sa présence ou de son absence).

« Enjeu 3 : Démocratie participative avec équilibre des pouvoirs (états, partenaires sociaux : entreprises, syndicats, etc.) »

(Table monde du travail, Musée de la civilisation, Québec, 6 avril 2018, scénario Une restructuration responsable).

L'urgence de la situation et la présence d'un certain déterminisme technologique ont été reconnues comme des facteurs pouvant nuire à la gouvernance participative. Le manque de temps évacuant la possibilité d'un travail démocratique doit être reconnu en connaissance de cause.

« Urgence au lieu de prendre le temps d'en faire un débat démocratique informé et participatif »

(Table monde du travail, SAT, Montréal, 13 mars 2018, scénario Une restructuration responsable).

JUSTICE SOCIALE

Les citoyens ont soulevé différents risques et enjeux liés aux biais que pourraient contenir les algorithmes, à l'accessibilité de l'IA, à la discrimination ou à l'exclusion de certains groupes d'individus qui pourraient en découler. Les conséquences de ces risques sur la diversité et l'équité sont des enjeux qui ont été considérés comme importants.

« Mettre en place des algorithmes non discriminatoires qui favorisent la diversité, l'inclusion et la justice sociale »

(Table monde du travail, Musée de la civilisation, Québec, 6 avril 2018, scénario L'IA comme passage obligé vers l'emploi).

Les enjeux relatifs à **l'accessibilité** sont ceux du comment garantir l'accès à l'IA et à ses usages. Ils sont associés au risque de restriction de l'accès à certains groupes de personnes ou catégories sociales. Les discussions ont également porté sur l'impartialité des systèmes algorithmiques et sur leurs biais discriminatoires potentiels, notamment ceux issus des données sur lesquelles l'algorithme est entraîné, ainsi que ceux issus de leur collecte, de leur analyse, ou encore du code lui-même.

« Les valeurs de justice (indépendance, impartialité, équité) prévalent sur les techniques lors du déploiement de ces outils. »

(Table système judiciaire et police prédictive, SAT, Montréal, 13 mars 2018, scénario Libération conditionnelle).

La **discrimination** qui pourrait apparaître si les deux premières catégories d'enjeux (accessibilité et exclusion) ne sont pas correctement considérées a été soulignée : il s'agit des effets discriminatoires des systèmes d'IA, que ce soit par le renforcement de discriminations actuelles (p. ex. liées au genre ou à la classe sociale) ou la création de nouvelles discriminations (p. ex. les personnes qui ne sont pas « connectées »). Les enjeux discriminatoires sont très proches du risque **d'exclusion** de certaines personnes, qu'elles refusent elles-mêmes de s'inclure dans la « société numérique » ou qu'elles soient exclues involontairement.

« Qu'est-ce qui arrive avec les gens qui n'ont pas de profil numérique ? Sont-ils désavantagés ? Doit-on confier 100% du recrutement à une IA ? Peut-elle saisir les critères d'embauche ? A-t-on le choix si tout le monde le fait ? Et la réputation numérique : comment l'évaluer ? »

Ces risques ont conduit les participants à identifier un enjeu de protection de :

1. La diversité ou la mixité des intelligences, des compétences, des individus et de la société dans son ensemble.

« Est-ce que l'IA reproduit simplement la même intelligence que celle que l'école forme ? Ne gagnerait-on pas à valoriser différentes formes d'intelligence ? »

2. L'équité pour que le fonctionnement de l'IA conduise à des décisions ou des recommandations.

« Partage des bénéfices de l'IA (gains de productivité). Équité entre les groupes sociaux, territoires (villes et régions), prise en compte des vulnérabilités. »

(Table monde du travail, Musée de la civilisation, Québec, 6 avril 2018, scénario Une restructuration responsable).

LIBERTÉS

Cette catégorie réfère aux enjeux relatifs au mantien des libertés individuelles, surtout en ce qui a trait à la liberté de choix — qu'il s'agisse de pouvoir suivre son propre choix face à une décision orientée par l'IA, mais également de pouvoir choisir de ne pas utiliser ces outils sans pour autant risquer une exclusion sociale (ce qui signifie que ces enjeux sont parfois très liés à la catégorie précédente).

AUTODÉTERMINATION

Des discussions relatives aux risques qu'une trop grande force déterministe soit attribuée aux systèmes algorithmiques ont eu lieu, notamment concernant les enjeux relatifs à la capacité d'autodétermination des individus (qui s'oppose à un risque de confiance aveugle en la technologie).

« Ce qui me touche le plus c'est que la grand-mère est exclue de la réflexion. L'infirmière robot d'accord, mais qu'est-ce que veut la grand-mère ? Il faudrait demander aux gens ce qu'ils veulent »

LIBERTÉ DE CHOIX

La possibilité d'expression d'un choix individuel ainsi que celle d'un droit de refuser d'utiliser une technologie ou de participer à un système de collecte de données ont également été discutées.

« Comment assurer que l'individu conserve sa liberté de choix et ne devienne pas dépendant de la technologie ? »

« Si on a besoin des données de tout le monde pour générer un bien-être collectif, est-ce qu'on doit obliger tout le monde à partager ses données ? Et si les personnes ne le font pas, quel sera alors cet effet sur le système ? C'est un choix de société à faire. »

UNIFORMISATION DE LA SOCIÉTÉ

L'uniformisation de la société concerne les enjeux relatifs aux risques qui accompagnent la tendance des IA à la catégorisation des individus au nom de la prédiction en matière de santé, d'éducation, de justice ou de mobilité. Cela pourrait amener à stigmatiser des individus et normaliser des comportements plutôt que d'en encourager la diversité.

« Risque d'avoir un profil type (normalisation des conduites) »

(Table ville intelligente et objets connectés, INM, 18 février 2018, scénario Réfrigérateur connecté).

MUTATIONS SOCIO-NUMÉRIQUES

Cette catégorie réfère aux discussions et enjeux relatifs aux transformations sociales et sociétales qui accompagnent ou accompagneraient le développement de l'IA. Ces transformations pourraient mener (ou non) à une véritable « transition numérique ».

ACCEPTABILITÉ

Les citoyens ont plusieurs fois soulevé les enjeux de l'acceptabilité de l'implémentation de l'IA et de l'adhésion sociale. Ces discussions ont porté sur les enjeux relatifs au maintien de la **confiance** du public dans la technologie (l'IA) et dans les différents secteurs qui seraient amenés à l'utiliser. Elles ont également porté sur les attentes projetées sur la technologie et sur les enjeux de la « technophobie ». Parfois, un certain **fatalisme** s'est dégagé, notamment concernant le déterminisme technologique et une certaine acceptation subie du développement de l'IA. La **légitimité** de l'utilisation de l'IA dans certains secteurs a parfois été remise en question.

« Maintenir et promouvoir la confiance de la population dans le système de justice »

(Table système judiciaire et police prédictive, Musée de la Civilisation, Québec, 6 avril 2018, scénario Libération conditionnelle).

COMPÉTENCES DE L'HUMAIN

Les participants ont plusieurs fois discuté des conséquences du développement de l'IA sur les compétences humaines. Par exemple, la **transformation des compétences humaines** a été abordée du point de vue des conséquences (surtout négatives) du développement de l'IA sur les capacités et connaissances.

« Crainte de dépassement de l'humain, capacité de l'humain à être à 360 ° (alors que l'IA a d'excellentes compétences très spécifiques). »

« Comment assurer que le dialogue avec le patient demeure (contact humain) et que le médecin ne perde pas son expertise et indépendance ? »

(Table santé, bibliothèque de Sainte-Julie, 25 mars 2018, scénario Hôpital intelligent).

Un risque de **dépendance à la technologie** (et plus particulièrement, ici, à l'utilisation de l'IA) a en effet été soulevé.

« On devient dépendant (de la technologie) »

« L'IA entraîne une dépendance dans la spécialisation et éloigne de la culture générale et l'autonomie d'apprentissage »

Les enjeux relatifs à la **littératie numérique** réfèrent à la nécessité de la formation de la population face aux pratiques et enjeux de l'IA, pour acquérir les compétences à la fois techniques et critiques nécessaires pour développer une capacité d'agir en tant que travailleur et citoyen dans une société numérique en transition.

« Pour garantir une utilisation avertie d'un dispositif comme AlterEgo, il est important que les jeunes, les parents et les enseignants soient sensibilisés à la façon dont les données collectées sont utilisées. Cela soulève un enjeu de connaissance impliqué par une démarche de littératie de l'IA. »

COMPÉTENCES DE L'IA

Concernant les compétences de l'IA, les enjeux relatifs aux **réels bénéfices** ont amené des discussions qui portent sur la remise en question des bénéfices ou des possibilités d'utilisation de l'IA envisagées.

« Comment assurer que nos outils IA respectent les principes fondamentaux de notre système de Justice ? »

(Table système judiciaire et police prédictive, Musée de la civilisation, Québec, 6 avril 2018, scénario Libération conditionnelle).

« Est-ce que l'IA remplit sa fonction d'améliorer et rendre accessible la santé et la qualité de vie des individus/communautés (rationalisation, déshumanisation du soin, effets inattendus et efficacité réelle des algorithmes, etc.) ? »

(Table santé, SAT, Montréal, 13 mars 2018, scénario Jumeaux numériques).

Assurer l'**efficacité et la validité** de l'IA, soit la pertinence de son utilisation et de ses compétences, a également été identifié comme un enjeu.

« Il faut garantir des recommandations de santé basées sur : 1) des algorithmes encadrés, validés, mis à jour (à partir du savoir scientifique) et intègres (sécurité/hacking) ; 2) des données complètes, vraies et non biaisées. »

(Table santé, bibliothèque Benny, Montréal, 18 mars 2018, scénario Jumeaux numériques).

« Si l'IA produit des conclusions fausses, comment va-t-on s'assurer d'évaluer sa performance ? Forcément, l'IA va évoluer, il faut prévoir certaines dispositions pour valider les résultats et prévoir une évaluation en continu. »

« Oui, il faut qu'il y ait a posteriori de chaque décision une évaluation de cette décision. Si on n'évalue pas la performance et les conséquences des jugements rendus par l'algorithme et qu'on continue d'utiliser l'algorithme, l'IA finit par se baser sur des erreurs. »

Le risque de remplacement de l'humain a également été abordé à plusieurs reprises, lié au rôle attribué à l'IA et les fonctions qu'elle pourrait remplir à la place de l'humain, aux avantages et inconvénients de son utilisation, à la manière de distribuer le système des compétences entre humain et IA.

« L'IA va compenser pour certains manques du système éducatif, mais est-ce la solution ? Le travail des professeurs sera réduit considérablement, ce qui permet un allègement, mais en même temps peut poser la question du remplacement »

Des discussions plus nuancées ont mis en évidence les enjeux du maintien d'un **équilibre entre les bénéfices et les risques** qu'apporteraient l'IA et ses compétences, ou sur la nécessité de tenir compte à la fois de ces bénéfices et de ces risques pour un développement responsable.

« Comment équilibrer l'implantation de l'IA dans les objets qui peupleront notre quotidien avec un développement harmonieux de la société (aspect culturel, bien-être, développement de l'enfant, candeur) et des êtres vivants ? » (Table ville intelligente et objets connectés, SAT, Montréal, 13 mars 2018, scénario Jouet intelligent).

COMPLÉMENTARITÉ HUMAIN — IA

Cette catégorie réfère aux discussions portant sur les avantages d'une complémentarité humain-IA ou sur les inconvénients d'une éventuelle « collaboration ». Est majoritairement ressortie la complémentarité entre, d'un côté, l'objectivité et la systématisation de l'IA et, de l'autre, la subjectivité et la contextualisation empathique de l'humain.

« Assurer la complémentarité IA-professeur en termes d'expertise et de relationnel avec les élèves. »

(Table éducation, SAT, Montréal, 13 mars 2018, scénario AlterEgo).

« Comment s'assurer que la décision en santé ne se base pas uniquement sur des données objectives et tienne compte du contexte, qu'elle prenne en considération le choix des utilisateurs ? »

(Table santé, bibliothèque du Boisé, Montréal, 17 mars 2018, scénario Assurance Santé).

« Justice objective de la prévision de l'IA versus intelligence subjective (basée sur l'expérience) »

RESPECT DE L'HUMAIN

Les citoyens ont soulevé les enjeux du respect de la nature et **condition humaine**. Ces discussions ont soulevé des réflexions sur ce qui définit un humain, ce qu'il va rester de l'humain, ou sur le comment conserver une primauté de l'humain dans le contexte du développement de l'IA et de la place qu'elle pourrait occuper.

« Qu'est-ce qu'un être humain ? Qu'est-ce qu'on préserve de l'humain ? Que veut-on préserver de l'humain ? »

Le risque de **déshumanisation** des activités et des services avec le développement de l'IA ou celui de l'apparition d'une nouvelle forme **d'isolement** — issue en particulier d'une diminution de la socialisation, ou de la délégation du lien social à des robots — ont été soulevés à plusieurs reprises.

« Il manque l'aspect humain au soin. Le rapport entre le(s) professionnel(s) de la santé et les patients »

« Comment assurer la dignité humaine et la place de l'être humain dans le système de justice ? »

(Table système judiciaire et police prédictive, Musée de la civilisation, Québec, 6 avril 2018, scénario Libération conditionnelle).

« Ça va faire une standardisation des causes aussi et on ne va pas assez prendre en compte la personne »

« Le relationnel avec l'IA au détriment des humains conduit à une solitude grandissante. »

TRANSFORMATION DES ACTIVITÉS

Cette catégorie réfère aux discussions relatives aux changements sociétaux qui accompagneraient le développement de l'IA et sur l'éventuelle transition numérique qui pourrait s'opérer dans les différents secteurs concernés et à différents niveaux (par exemple, l'IA transforme la connaissance, la ville, la conception du travail, etc.)

« On rationalise la santé »

« Rédéfinition/transformation de la nature de la relation entre l'enseignant et les étudiants dans l'espace pédagogique et la modification des rapports au savoir »

(Table éducation, SAT, Montréal, 13 mars 2018, scénario Nao).

« L'augmentation des capacités mentales par le transhumanisme va-t-il rendre l'éducation vétuste ? »

« Il y a un risque de cristallisation du droit. Plus l'IA rend de décisions dans un sens donné plus elle sera portée à rendre des décisions dans ce même sens. »

« Dans 30 ans, les gens vont dormir, travailler, etc. dans leur voiture, et celle-ci ce ne sera plus simplement un dispositif dédié au déplacement. La mobilité aura un autre sens. »

VIE PRIVÉE ET CONFIDENTIALITÉ

ANONYMAT, CONFIDENTIALITÉ ET DILEMME AVEC LA BIENFAISANCE

Cette catégorie réfère aux enjeux relatifs au respect de **l'anonymat** et de la **confidentialité**. Les discussions ont porté sur la réelle possibilité de respecter cet anonymat avec le développement responsable de l'IA, ou sur le comment garantir la protection de la confidentialité de certaines données « sensibles », voire de restreindre leur accès à certaines personnes et usages qui seraient plus justifiés que d'autres. L'IA a été présentée tantôt comme le problème, tantôt comme la solution à ce genre d'enjeux. Un **dilemme** a été soulevé à plusieurs reprises, en particulier dans le domaine de la santé. Ce dilemme réfère à l'opposition entre la bienfaisance (qui supposerait de collecter un maximum de données et pas seulement des données objectivables afin de garantir une approche plus humaine et plus contextuelle de l'IA) et le respect de la vie privée et de la confidentialité (qui serait défié par cette même collecte).

« La confidentialité n'existe plus, c'est un mythe. On a tenté l'anonymisation des données, ça ne marche pas. À présent on peut imposer que seuls les algorithmes voient les données, pas les acteurs humains qui les manipulent »

DROIT À L'OUBLI

Les discussions ont également porté sur la création d'un droit à l'oubli (avoir la possibilité d'effacer des données personnelles) et des enjeux et conséquences de sa mise en place.

« Droit à l'oubli (durée de rétention), droit à la modification, droit à la suppression »

(Table système judiciaire et police prédictive, bibliothèque Père-Ambroise, Montréal, 3 mars 2018, scénario Arrestation préventive).

INTRUSION

Des discussions qui portent sur les risques d'intrusion dans la vie privée des individus, le non-respect de la vie privée et aux moyens de garantir cette protection ont eu lieu à plusieurs reprises.

« Comment assurer le respect des différentes composantes de la vie privée (oubli, propriété, consentement, portabilité) dans le contexte de l'utilisation des objets connectés ? »

(Table ville intelligente et objets connectés, SAT, Montréal, 13 mars 2018, scénario Jouet intelligent).

PROPRIÉTÉ DES DONNÉES PERSONNELLES

Cette catégorie renvoie aux enjeux relatifs à la propriété des données personnelles, à leur définition, aux conséquences de cette propriété sur le respect de la vie privée (dans quelle mesure un individu est et restera propriétaire de ses propres données ?) et à la protection de la « réputation numérique » des individus.

« Les données relatives à la vie privée devraient être la propriété des personnes concernées et partagées selon des règles votées démocratiquement. »

(Table santé, INM, Montréal, 18 février 2018, scénario Jumeaux numériques).

SURVEILLANCE

Les enjeux relatifs à la surveillance sont liés à l'accessibilité des données et au profilage qui entraîne des préoccupations relatives à la surveillance de masse (et continue) des individus risquant ainsi de porter atteinte à la fois à la vie privée et à aux libertés individuelles.

« Comment vivre une vie saine quand on est constamment surveillé ? »

« Va-t-on être capable de retracer tous les déplacements des gens ? »

« Est-ce qu'une instance supérieure, gouvernement ou entreprise, pourrait prendre le contrôle de mon véhicule ? »

CONSENTEMENT LIBRE ET ÉCLAIRÉ

Les discussions ont également porté sur la capacité à consentir aux usages de l'IA et des données personnelles.

CONSENTEMENT LIBRE

Est remise en question ici la réelle indépendance des individus concernant le choix de partager ou non leurs données (personnelles), d'avoir un réel impact sur leur gestion ou de choisir les fins pour lesquelles elles seront réutilisées.

« Est-ce qu'on est vraiment libre de ne pas partager ses données ? »

« Si on partage publiquement, est-ce qu'on consent vraiment à la réutilisation ? »

CONSENTEMENT ÉCLAIRÉ

L'enjeu est ici lié aux modalités d'information nécessaires aux individus pour qu'ils puissent consentir de manière éclairée, il touche l'accès à l'information et la compréhension de cette information. Cet enjeu est très lié au niveau de littératie numérique des citoyens et aux enjeux de transparence.

« Au carrefour de ces enjeux liés aux données collectées et interprétées, et à celui de l'autonomie de l'élève, il y a la question du consentement éclairé (des enfants et des parents) »

ENVIRONNEMENT-ÉCOLOGIE

On touche ici aux enjeux concernant l'impact du développement et de l'utilisation responsable de l'IA sur l'environnement ainsi que son coût énergétique.

« On oublie de parler de l'aspect environnemental : le stockage des données, le problème d'une accumulation outrancière des données et des coûts énergétiques que cela implique. »

INFLUENCES

Ces enjeux réfèrent aux inquiétudes concernant les influences (qu'elles soient indues ou non), voire les manipulations potentielles issues de l'utilisation de l'IA. Pour entretenir une certaine liberté dans les choix orientés par l'IA et éviter d'accorder une confiance aveugle à ces dispositifs, les citoyens ont ici reconnu la nécessité de cultiver une pensée critique pour tous ceux qui interagissent avec l'IA.

LOBBYISME

Les citoyens s'inquiètent de l'apparition d'une nouvelle classe de lobbys avec le développement de l'IA qui pourrait parfois avoir trop de pouvoir et d'influence sur le système de santé, les objets connectés ou les véhicules autonomes.

« Est-ce qu'il revient au politique de déterminer quel algorithme sera utilisé ? Qu'en est-il des lobbys des créateurs d'algorithmes ? »

MANIPULATION

Les participants s'inquiètent d'un risque de manipulation qui plane sur les utilisateurs à mesure que leurs actes et décisions sont de plus en plus influencés par des mécanismes d'IA, que ce soit à leur insu ou via des incitatifs plus explicites.

« Jusqu'où la machine peut-elle nous influencer dans nos décisions ? Sait-on à quel point les suggestions du frigo connecté influenceront notre quotidien ? »

« Influence insidieuse sur nos comportements sans qu'on nous l'ait demandé et qu'on l'ait accepté »

« Risques d'influence : Comment rendre visibles les risques d'influences (consommation, discernement) liés à l'usage des objets connectés ? Comment assurer le respect des intérêts de chacun (consommateur et citoyen, mais aussi compagnies) ? Qui et comment décide-t-on des lignes directrices pour développer ces (éco) systèmes ? »

(Table ville intelligente et objets connectés, SAT, Montréal, 13 mars 2018, scénario Jouet intelligent).

PATERNALISME

L'exposition à différentes formes de paternalisme et de contrôle (des entreprises, de l'état) a été mentionnée à plusieurs reprises. Celle-ci pourrait être accentuée par des systèmes incitatifs, mais également par une dépersonnalisation des relations (notamment la relation de soin).

VULNÉRABILITÉ DES PERSONNES

Les citoyens ont reconnu que toutes les personnes n'avaient pas le même niveau de vulnérabilité face aux risques d'influence présentés. Leur protection particulière a été soulignée comme un enjeu important.

PARTAGE DE LA RESPONSABILITÉ

Cette catégorie réfère aux enjeux du partage des responsabilités face aux risques du développement responsable de l'IA et aux conséquences des décisions.

DÉRESPONSABILISATION

La déresponsabilisation fait ici référence aux inquiétudes relatives au risque de déresponsabilisation face au développement de l'IA qui pourrait s'opérer par une délégation de cette responsabilité aux algorithmes (considérant leur autonomie croissante ou la perception d'une autonomie croissante).

« Risque de déresponsabilisation de l'enseignant qui se plierait au "syndrome du diagnostic" combiné au risque de renforcer un certain profil de l'élève. »

(Table éducation, bibliothèque Père-Ambroise, Montréal, 3 mars 2018, scénario AlterEgo).

« Ça crée une déresponsabilisation ; supposons que je sois hyperactif, la machine le confirme, donc je fais moins d'efforts. Mais faut que tu fasses partie de la solution mon chum. La façon de travailler va changer. Il va y avoir une modification de la tâche de l'enseignant, c'est certain. »

« La connaissance est liée à la responsabilité. Il y a un risque de déresponsabilisation s'il y a une perte de connaissance. Une perte d'esprit critique de la part des juges et des personnes. »

« Comment assurer que l'IA reste un service et que les différents acteurs (individus, programmeurs, société, etc.) ne soient pas déresponsabilisés, restent vigilants et que l'individu soit toujours en contrôle ? »

(Table ville intelligente et objets connectés, Musée de la civilisation, Québec, 6 avril 2018, scénario Réfrigérateur connecté et Empreinte Carbone).

IMPUTABILITÉ

Cet enjeu réfère à l'identification de qui est responsable ou imputable dans différentes situations liées au développement de l'IA (l'utilisateur, le développeur, l'algorithme, etc.?).

« Qui détient les données d'apprentissage, qui les utilise, pendant combien de temps ? Qui les protège ? »

(Table éducation, bibliothèque de Sainte-Julie, 25 mars 2018, scénario Nao).

« Qui pilote tout ça ? Quels pouvoirs a l'organisation ou la compagnie sur cet outil ? Sera-t-on dépendants de cette compagnie-là ? Si ça devient une priorité nationale, quels choix vont être faits pour les programmes pédagogiques à partir de l'implantation ? Est-ce que c'est public ? Privé ? C'est tout l'écosystème de l'éducation qui est redessiné. »

« Qui gère l'algorithme, qui le contrôle, qui surveille celui qui le programme ? »

RESPONSABILITÉ PARTAGÉE

Les discussions ont également porté sur la délimitation des responsabilités face au développement de l'IA, la complexité de ce partage et la nécessité de tenir compte de la pluralité des responsabilités et des acteurs.

« L'enjeu relatif aux responsabilités individuelles et partagées, et possiblement conflictuelles, des différents acteurs (gouvernements, professionnels de la santé, patients, entreprises privées, chercheurs et gestionnaires, etc.). »

(Table santé, SAT, Montréal, 13 mars 2018, scénario Vigilo).

« Enjeu 2 : Circonscrire les rôles et les responsabilités de chacun (institutions, étudiants, professeurs) afin d'encadrer l'introduction de l'I.A.»

(Table éducation, SAT, Montréal, 13 mars 2018, scénario Nao).

« Je ne connais pas de profs qui se déresponsabilisent vis-à-vis de leurs élèves. Mais il faut impliquer le plus de monde possible, faire du multidisciplinaire. Ne pas faire du prof le premier et le dernier responsable de l'IA ou des diagnostics de l'IA. S'assurer que l'utilisation pédagogique de l'IA soit une responsabilité partagée. »

SOUVERAINETÉ DE LA DÉCISION

Les enjeux relatifs à la souveraineté de la décision font écho aux attentes normatives précisées dans les recommandations (« Grandes directions attendues ») qui mentionnent que l'IA doit rester un outil, un assistant ou une ressource supplémentaire qui apporte une information additionnelle. Ces recommandations ont fait suite aux discussions portant sur les enjeux de la souveraineté de la décision, soit qui de l'humain ou de l'IA a le dernier mot.

« Les algorithmes devraient toujours être consultatifs et non décisionnels. L'absence de modération humaine est problématique, l'algorithme ne tenant pas compte de tous les aspects de l'individu. »

(Table santé, bibliothèque Père Ambroise, Montréal, 3 mars 2018, scénario Jumeaux numériques).

« Le problème avec l'interprétation des diagnostics d'AlterEgo c'est qu'il ne faut pas oublier que l'intervention humaine est nécessaire. On ne peut pas se fier uniquement à la machine. »

« On délègue beaucoup de micros décisions à des IA et systèmes interconnectés au détriment de l'humain ».

STRESS – ALARMISME – ANGOISSE

Les participants s'inquiètent que le développement de l'IA génère du stress, de l'angoisse ou de l'anxiété issus notamment d'un surplus d'informations et de notifications ou d'une absence de contact humain.

« Comment les élèves vont-ils développer leur autonomie académique et apprendre à gérer leur stress et leurs émotions quand ils n'auront pas accès à AlterEgo pendant leur enseignement supérieur ? »

(Table éducation, bibliothèque Benny, Montréal, 18 mars 2018, scénario AlterEgo).

« Il faut garantir le bien-être de l'individu lorsqu'on le soigne, l'informe : ne pas être alarmiste. »

(Table santé, bibliothèque Benny, 18 mars 2018, scénario Jumeaux numériques).

SÉCURITÉ ET INTÉGRITÉ DES SYSTÈMES

Les enjeux liés à la **fiabilité** des systèmes (IA et données) ont été discutés à différents niveaux : validité, infaillibilité et robustesse, intégrité des systèmes et des personnes qui les gèrent. Ont également été discutées la vulnérabilité de ces systèmes (bogues, erreurs, etc.) et les conséquences des failles dans les différents paramètres de fiabilité. Les enjeux relatifs au risque de **panne** des systèmes et à la gestion de ce risque ont également été soulevés. Ces enjeux sont étroitement liés aux biais et aux compétences de l'IA. Les citoyens s'inquiètent des risques de **piratage ou sabotage** des données collectées et des algorithmes, qu'ils soient intentionnels ou non, et des risques associés aux possibles **usages détournés** des utilisations initialement prévues des données et algorithmes (sans qu'il s'agisse pour autant de piratage) et qui seraient problématiques.

« Je ne veux pas que l'on me juge plus tard de choses faites dans le passé »

« Et si un hacker prenait le contrôle du développement pédagogique de certains élèves ? Ou si les parents pouvaient avoir aussi beaucoup plus d'impact sur les performances scolaires de leurs enfants ? Le hacker ou les parents pourraient choisir le contenu, alors comment AlterEgo analyserait les données. Par exemple, des parents qui veulent éviter que leur enfant s'investisse dans une carrière artistique pourraient détourner AlterEgo pour servir cet intérêt. »

L'intention et la malveillance dans l'utilisation problématique ou non sécuritaire de l'IA ont été identifiées comme des paramètres importants. Les citoyens ont mis l'accent sur la difficulté de différencier un acte malveillant d'un acte problématique issu d'une bonne intention et sur les conséquences de cette différenciation.

« Même avec de bonnes intentions, on peut causer des problèmes (modèle inexact) »

« Comment différencier un comportement passager sans intention de nuire versus une réelle prise de décision avec intention de commettre un crime ? »

Plusieurs fois, les discussions ont tourné autour de la possibilité d'atteindre le risque zéro et on s'y est demandé s'il est souhaitable de le faire.

« L'objectif 0 accident doit-il être atteint à n'importe quel prix ? Est-ce que cet objectif en vaut vraiment la peine ? »

Concernant la protection de la sécurité, plusieurs dilemmes ont été identifiés lors des discussions :

> Avec la transparence (garantir la transparence pourrait augmenter les risques de piratage) ;

> Avec l'efficacité (assurer la plus grande sécurité possible implique un compromis avec l'efficacité du système, qui doit être sécuritaire sans pour autant devenir inopérant) ;

> Avec le respect des libertés individuelles et de la vie privée (dans le cas particulier d'arrestations préventives, qui impose la surveillance au nom de la sécurité publique).

TRANSPARENCE

L'enjeu de transparence s'est formulé autour de la capacité de comprendre une décision algorithmique et d'agir face à elle, que ce soit en tant que citoyen dans sa vie quotidienne ou en tant que professionnel ayant recours à l'IA dans le cadre de l'exercice de ses fonctions.

EXPLICABILITÉ ET COMPRÉHENSION

Ces enjeux sont relatifs à l'explicabilité de la décision et à la « boîte noire », à l'importance de rendre compte des processus qui mènent l'IA à un résultat ou à l'intelligibilité de l'information et l'importance de la vulgarisation.

« Transparence des variables utilisées, des données, des paramètres. Expliquer une décision en langage naturel. »

(Table monde du travail, bibliothèque Mordecai-Richler, Montréal, 10 mars 2018, scénario L'IA comme passage obligé vers l'emploi).

« La complexité du monde des algorithmes ne permet pas de comprendre comment l'IA a procedé. (...) on ne demande pas autant de transparence de la part des juges, donc pourquoi en demander autant pour l'algorithme ? »

DROIT À L'INFORMATION VERSUS DROIT DE NE PAS SAVOIR.

Ce dilemme a particulièrement été observé dans le secteur de la santé et oppose le droit de ne pas savoir (l'ensemble des prédictions diagnostiques issues d'une IA par exemple) au droit de savoir (afin de respecter l'autonomie du patient et son consentement). Le droit de ne pas savoir pourrait se justifier au regard de la bienfaisance (si certaines recommandations sont alarmistes et peu certaines).

5. PISTES DE SOLUTION ET D'ENCADREMENT POUR UN DÉVELOPPEMENT RESPONSABLE DE L'IA

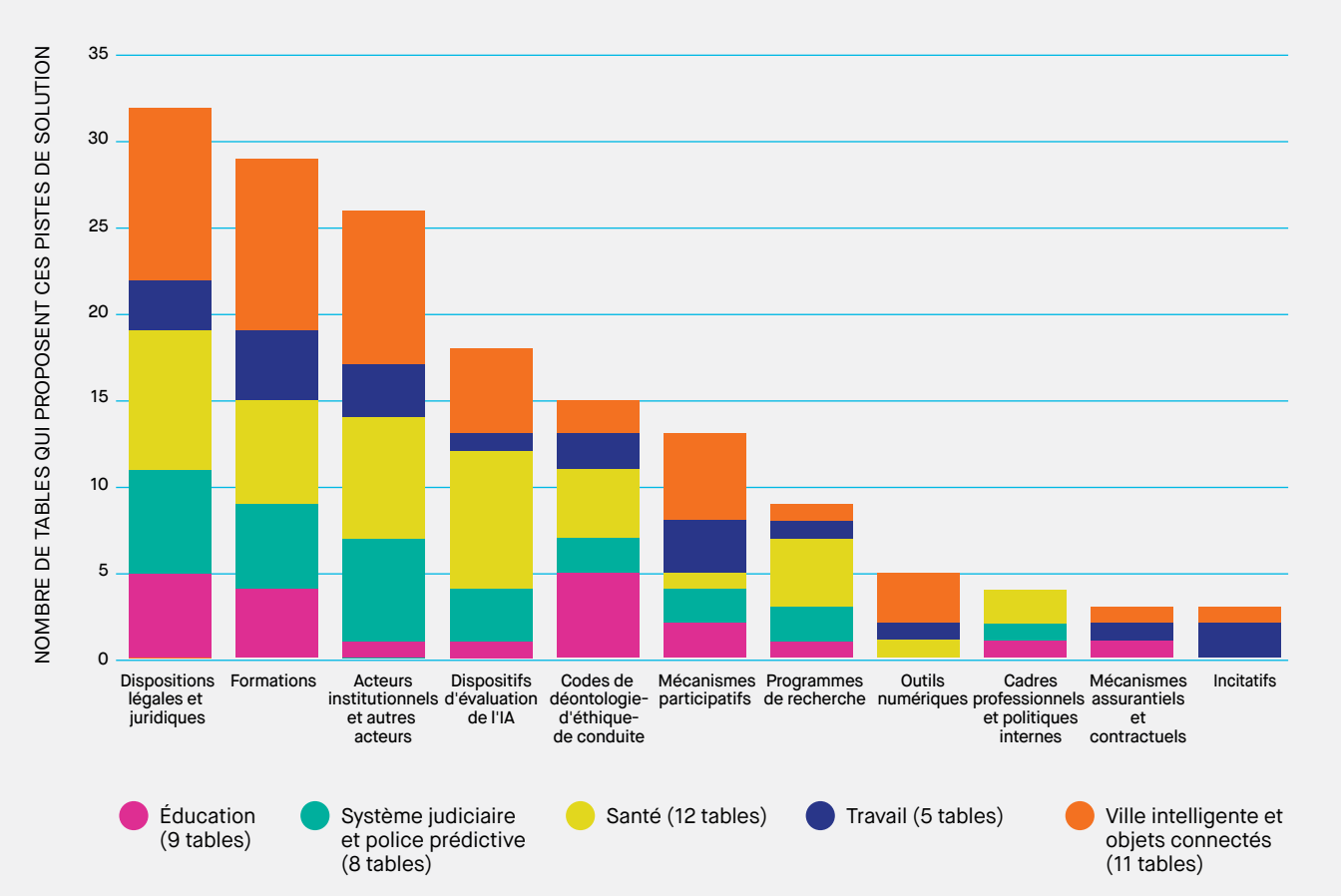
5.1.

INTRODUCTION

Les citoyens ayant participé aux journées de coconstruction ont été invités à proposer des pistes de solution en réponse aux enjeux précédemment identifiés. Un total de 190 pistes de solutions ont été formulées et adoptées de manière consensuelle lors de ces activités (bien que d'autres propositions aient pu être discutées autour des tables). Sont entendus ici par pistes de solution les mécanismes concrets envisagés par les citoyens pour répondre aux enjeux précédemment identifiés.

Seules les pistes de solutions et d'encadrement formulées sur les affiches ont pu être comptabilisées. Cependant, d'autres recommandations ont été discutées ou proposées (lors de la rédaction des *Unes* ou des discussions). Pour des raisons de cohérence et de faisabilité, ces dernières n'ont pu être comptabilisées dans le nombre total de recommandations, mais ont été considérées et analysées pour la rédaction de la présente section.

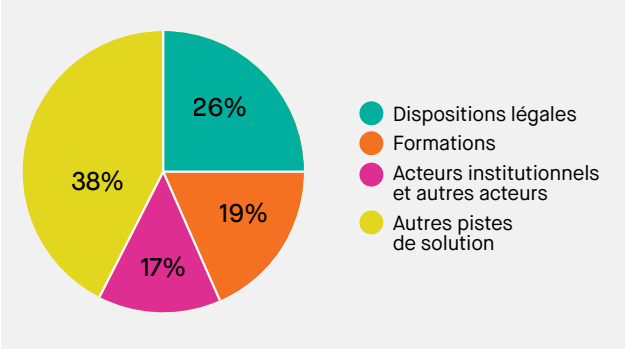
Tableau 1 : Les pistes de solution proposées pour répondre aux enjeux identifiés



Toutes les tables de coconstruction se sont entendues sur 3 grandes pistes de solution pour garantir un développement socialement responsable de l'IA, et ce, quel que soit le secteur :

1. Des dispositions légales et juridiques
2. La mise en place de formations pour tous
3. L'identification d'acteurs clés et indépendants pour la gestion de l'IA

Tableau 4 : Les trois principale pistes de solutions proposées par les tables de coconstruction



Quel que soit le secteur, l'ensemble des tables s'est entendu pour recommander la mise en place de dispositions légales et juridiques adaptées à la réalité des enjeux liés au développement de l'IA et à la gestion des données personnelles (en particulier, les données massives). Ces dispositions contraignantes réfèrent toutes au droit ou à la loi. Elles peuvent prendre la forme de lois et règlements ; de la défense de nouveaux droits fondamentaux ou encore de politiques publiques (allant de la mise en place de programmes sociaux et de charte à la création d'une citoyenneté numérique).

La mise en place de formations accessibles à tous a également été largement recommandée, tant pour les professionnels des secteurs concernés (afin de garantir une utilisation adéquate des systèmes d'IA dans leur pratique) que pour l'ensemble de la population (afin de garantir que tous puissent participer au débat et acquérir un niveau de base en littératie numérique).

Les citoyens ont également identifié des acteurs institutionnels et des acteurs clés (existants ou à créer) indépendants et compétents, qui seraient garants du développement responsable de l'IA. Les acteurs nommés sont des personnes (ex. ombudsman, commissaire aux comptes, commissaire à la vie et au bien-être) ou des groupes de personnes (ex. la mise en place d'un centre de l'intelligence artificielle pour la sécurité civile, d'une ligne 1-800 contre la discrimination par des objets connectés ou d'un ministère de l'éthique des données et de la protection numérique).

En recommandant majoritairement ces trois mécanismes comme pistes de solution, une nette tendance se dégage concernant la position des citoyens québécois ayant participé aux activités sur le mode de gouvernance de l'IA : celle-ci devrait être plutôt Étatique. En effet, la mise en place d'incitatifs pour les entreprises ou de mécanismes assurantiels et contractuels qui correspondent à une gestion plus libérale sont les pistes de solution les moins recommandées. Ces recommandations n'en sont pas moins cohérentes et instructives. Le développement d'incitatifs — qui vise à encourager le développement responsable — a fait consensus à différentes tables, comme la mise en place de *Quotas diversité* (qui récompenseraient les entreprises qui garantissent de ne pas exclure ou discriminer certaines minorités par le biais de leur IA) ou des subventions aux entreprises qui mettent en place des transitions pour les employés qui voient leur travail se faire remplacer par l'IA. La création de contrats entre les différentes parties prenantes du développement de l'IA et les utilisateurs, ou de mécanismes assurantiels garantissant la protection des individus face au développement de l'IA a également été proposée.

Pour tous les secteurs, les citoyens ont proposé la création de mécanismes d'évaluation technique et éthique de l'IA. Notamment, la mise en place d'un système de certification (ou label) comme garantie éthique a été mentionnée à de nombreuses reprises. Différentes tables ont également recommandé la mise en place de codes de déontologie ou d'éthique (qu'il s'agisse d'adapter les codes existants ou d'en créer des nouveaux) et de mécanismes participatifs (ex. coconstructions ou consultations publiques) afin de garantir un développement démocratique

de l'IA et de sa gestion. La mise en place de cadres professionnels (et différentes procédures internes aux entreprises ou institutions) qui ne sont pas des codes déontologiques a également été discutée.

L'importance de la mise en place de programmes de recherche dans des disciplines variées (ex. philosophie, sciences sociales, bioéthique) favorisant le développement de nouvelles connaissances et la création d'outils numériques (ex. formulaire de consentement électronique et interactif dans le secteur de la santé, dossier numérique individuel dans le secteur monde du travail) ont également été mentionnées.

Les sections subséquentes présentent l'ensemble des pistes de solution et d'encadrement formulées par les citoyens en fonction des secteurs d'application de l'IA. Ces pistes de solutions et d'encadrement, précisées en mécanismes concrets, n'ont pas toutes été discutées et développées avec le même niveau de précision. S'il est évident qu'il est difficilement envisageable de mettre en place toutes ces recommandations, considérant leur variété et le fait qu'elles sont parfois contradictoires, une présentation exhaustive offre cependant une vision d'ensemble particulièrement riche de la diversité des solutions envisagées par les citoyens en termes de gestion de l'IA.

5.2. ÉDUCATION

Tableau 5 : Pistes de solution ou grandes directions pour le secteur de l'éducation

	Nombre de pistes de solutions formulées
Dispositions légales	8
Formations	7
Codes de déontologie-d'éthique-de conduite	5
Mécanismes participatifs	2
Acteurs institutionnels et autres acteurs	1
Dispositifs d'évaluation de l'IA	1
Programmes de recherche	1
Cadres professionnels et politiques internes	1
Mécanismes assurantiels et contractuels	1
Total	27

DISPOSITIONS LÉGALES ET JURIDIQUES

La nécessité de créer et de renforcer certaines législations a été soulevée par les participants concernant le développement de l'IA dans le secteur de l'éducation. Par exemple en ce qui concerne l'utilisation des données, ont été recommandés un droit à l'oubli ; une « date d'expiration » ; et aucun partage par défaut avec d'autres services sans motif sérieux. Le droit à l'oubli a souvent été identifié comme le besoin de créer une « *politique de destruction des données* » pour préserver la faculté des élèves de refaçonner leur identité et de s'améliorer. La nécessité de renforcer la protection de la vie privée (notamment, relative aux données des jeunes) et la transparence en ce qui concerne la collecte des données (notamment en favorisant des formats facilement compréhensibles pour les utilisateurs) a également été soulevée. Pour les participants, un encadrement légal où

« en aucun cas, l'utilisation de l'intelligence artificielle ne devrait limiter les possibilités d'avenir de l'utilisateur, que ce soit socialement, économiquement, etc »

(Table de l'INM, Montréal, 18 février 2018, Scénario AlterEgo), devrait être mis en place.

D'autres initiatives ont aussi été formulées comme la création d'un règlement pour que les parents et les élèves puissent choisir d'utiliser ou non des dispositifs d'IA, de baliser l'implication de l'industrie dans le système éducatif pour veiller à une utilisation éthique de l'IA et enfin de prévoir des stratégies (via des politiques publiques) qui éviteraient le « hacking pédagogique » en conservant les données de manière cryptée.

D'autre part, certains citoyens ont développé l'idée de créer une loi ou un règlement qui viserait à

« développer un langage commun (inspiré de la santé avec le détail des étiquettes alimentaires sur les aliments transformés) pour réduire le fossé entre la technologie et les utilisateurs »

(Table de la bibliothèque de Laval, 24 avril 2018, scénario Nao).

FORMATIONS

En rapport à l'éducation, les participants ont reconnu la nécessité d'être proactif dans la mise en place de formations pour toute la communauté concernée par le développement de l'IA dans ce secteur. Cette formation devrait porter sur la littératie numérique, la littératie médiatique, mais également l'éthique et les enjeux liés à l'intégration de l'IA dans le milieu éducatif. Cette formation pourrait par exemple prendre la forme d'un accompagnement en littératie numérique pour les parents comme pour les élèves ou s'intégrer directement dans la formation initiale des citoyens.

Les citoyens ont également recommandé de former plus spécifiquement les professionnels du système éducatif, en incluant par exemple le développement de compétences de travail « en équipe » avec les dispositifs d'IA à travers le cursus de formation initiale et universitaire des professeurs (ex. une certification au B.Éd. ou un système d'accréditation). Cette formation devrait être à la fois technologique (comment utiliser l'IA), mais également portée sur les techniques d'enseignement avec l'IA (comment s'organisent les séquences d'enseignement et insister sur le fait que ce sont les professionnels du milieu qui orchestrent l'IA et non l'inverse).

« Accréditer des agents (autant psychoéducation, qu'enseignants déjà actifs) de changement par institution d'enseignement pour intégrer graduellement l'IA en milieu scolaire »

(Table de la SAT, Montréal, 13 mars 2018, scénario AlterEgo).

L'importance de la mise en place de formations adéquates a ainsi été soulevée. Elles auraient pour objectif de fournir les informations appropriées

permettant aux parties prenantes d'assumer leur responsabilité face à l'IA, de façon à éviter que les enseignants accordent une confiance aveugle aux dispositifs d'IA en éducation. Ces formations accélèreraient la compréhension des acteurs dans le milieu de l'apprentissage et favoriseraient leur mobilisation pour développer l'IA au service de l'autonomie des apprenants tout en les préparant à composer avec ces réalités. Ces formations devraient permettre de développer les compétences humaines, donner un certain pouvoir d'action pour encadrer, voire redéfinir le développement à venir de l'IA.

« Sensibiliser à l'utilisation responsable de l'I.A. et valoriser une pluralité de rapport au savoir »

(Table de la SAT, Montréal, 13 mars 2018, scénario Nao).

CODES DE DÉONTOLOGIE-D'ÉTHIQUE-DE CONDUITE

Les citoyens ont également recommandé la mise en place de codes de déontologie ou d'éthique professionnelle pour les enseignants ; qui se pencheraient, entre autres, sur différents principes d'éthique (p. ex. justice) pour l'utilisation de l'IA en milieu éducatif. Ces codes permettraient d'encadrer la profession pour éviter le désengagement du professeur ainsi que les risques d'abus, de profilage ou de discrimination.

« S'assurer que l'utilisation de l'IA soit une responsabilité pédagogique partagée (personne de soutien, famille, enseignants, robot) »

(Table de la SAT, Montréal, 13 mars 2018, Scénario Alter Égo).

« Exercer la profession d'enseignant dans le respect de la qualité relationnelle et émotionnelle de l'humain. »

(Table de la bibliothèque de Sainte-Julie, 25 mars 2018, scénario Nao).

MÉCANISMES PARTICIPATIFS

Les citoyens ont proposé la mise en place de Communautés IA open source dans les bibliothèques publiques, afin d'ouvrir la « boîte noire » de l'IA. L'idée de mener des États généraux par le biais de consultations dans le milieu éducatif sur le développement socialement responsable de l'IA en éducation a également été proposée.

« Consultation dans le milieu éducatif pour faire un État des lieux et définir les rôles et responsabilités de chacun des acteurs »

(Table de la SAT, Montréal, 13 mars 2018, scénario Nao).

ACTEURS INSTITUTIONNELS ET AUTRES ACTEURS

Il a été proposé de créer un Comité permanent québécois multipartite qui ne serait pas uniquement composé de fonctionnaires ministériels, mais aussi de représentants de parents, d'élèves, de professeurs, de bibliothécaires et de chercheurs. Il s'agirait d'un lieu de débat public et de contre-pouvoir aux compagnies privées. Ce comité aurait pour mandat de conseiller le gouvernement (recommandations exécutoires) ; préparer des codes de déontologie et formations ; veiller à l'existence de licences open source et de consulter les citoyens. Il a aussi été recommandé de mettre en place des comités d'éthique qui réaliseraient un processus de consultation à tous les stades de l'évolution de la technologie, tout en vérifiant sa bonne acceptabilité sociale. L'idée de la création d'un comité paritaire, inclusif et diversifié composé d'acteurs du milieu de l'éducation a été proposée. La responsabilité de la création de ce comité devrait, selon les citoyens, être confiée à un ministère. Enfin, certains participants ont proposé la création d'un

« Ministère de l'intégration et de l'accès technologique pour créer des formations et des certifications »

(Table de la bibliothèque de Laval, 24 mars 2018, scénario Nao)

DISPOSITIFS D'ÉVALUATION DE L'IA

Pour les participants, il est de mise de créer des certifications, notamment pour garantir le respect de certains standards à savoir : le respect, le choix conscient et la liberté. Aussi, certaines certifications pourraient offrir la garantie que les algorithmes ne soient pas utilisés pour remplacer les professeurs. Les participants ont préconisé des tests et des observations en classe afin de s'assurer que ce type d'outils ne dérange pas les étudiants.

PROGRAMMES DE RECHERCHE

Les citoyens ont recommandé le développement conjoint ou parallèle de la technologie et de la créativité humaine par le biais de programmes de recherche qui doivent être menés par des acteurs interdisciplinaires. Ces programmes pourraient porter par exemple sur la technologie et la santé mentale ; sur comment assurer la liberté de choix de l'utilisation de l'IA ; ou comment protéger l'autonomie humaine dans la décision. Ils ont aussi reconnu la nécessité de mettre l'IA au service de la recherche en éducation, afin d'agir le plus tôt possible dans l'apprentissage de l'enfant.

CADRES PROFESSIONNELS ET POLITIQUES INTERNES

Si les milieux scolaires intègrent l'IA, les citoyens étaient d'avis que cela devrait être fait de manière responsable. Pour ça, ils ont recommandé deux pistes de solutions : l'instauration d'incitatifs qui viseraient à encourager

« les établissements scolaires à se doter de politiques internes pour encadrer l'intégration de l'IA »

(Table de la SAT, Montréal, 13 mars 2018, Scénario Nao)

ou encore d'établir des protocoles ou des cadres qui permettraient d'identifier certains repères pour faciliter l'intégration responsable de l'IA en contexte scolaire.

MÉCANISMES ASSURANTIELS ET CONTRACTUELS

Pour préserver le bien-être des élèves, les citoyens ont indiqué qu'un engagement clair devrait être émis. Cet engagement pourrait prendre la forme d'un « contrat social ou moral » qui devrait être signé par toutes les parties prenantes. Sa mise en place permettrait de

« bien comprendre le degré de responsabilité pour préserver le bien-être de l'élève »

(Table du Musée de la civilisation, Québec, 7 avril 2018, scénario AlterEgo), mais aussi de permettre le droit de retrait des professeurs.

5.3.

SYSTÈME JUDICIAIRE ET POLICE PRÉDICTIVE

Tableau 6 : Pistes de solution ou grandes directions pour le secteur de du système judiciaire et de la police prédictive

	Nombre de pistes de solutions formulées
Dispositions légales	9
Acteurs institutionnels et autres acteurs	7
Dispositifs d'évaluation de l'IA	5
Formations	5
Codes de déontologie-d'éthique-de conduite	2
Mécanismes participatifs	2
Programmes de recherche	2
Cadres professionnels et politiques internes	1
Mécanismes assurantiels et contractuels	1
Total	34

DISPOSITIONS LÉGALES ET JURIDIQUES

En ce qui concerne le système judiciaire et la police prédictive, l'impératif est donné à l'établissement de lois et règlements sur la transparence : il s'agit d'exiger de la transparence de la part des compagnies privées et publiques collectant des données criminelles, et également de rendre explicites et interprétables les processus de décisions quand celles-ci sont rendues par des algorithmes. L'explicitation de la décision doit s'accompagner de mesures permettant d'avoir accès aux algorithmes mobilisés et être intelligible. Comme premier mécanisme de transparence, plusieurs tables ont proposé que les IA utilisées dans le secteur judiciaire — voire toutes les IA du secteur public — soient développées sous licence libre et en code ouvert. En termes de droit, il s'agit de garantir « le droit à une défense pleine et entière », notamment en ayant la possibilité de contester une décision en soulevant des vices de fond ou de forme (Table du Musée de la civilisation, Québec, 6 avril 2018, scénario Libération conditionnelle).

Cet impératif de transparence va de pair avec la mise en place de dispositions légales donnant le droit, estimé comme fondamental, d'être jugé par un être humain pour préserver une justice procédurale et une individualisation de la peine, mais aussi que la procédure d'appel d'une décision informatique soit

toujours entendue par un juge humain. Soulignant la nécessité du droit de s'adapter à une nouvelle réalité technologique avec l'IA dans la prise de décision en matière de justice, plusieurs débats ont eu lieu autour de la conciliation entre agents humains et artificiels dans ce processus. Le consensus se formule ainsi :

> « Droit d'appel devant un juge humain : La procédure d'appel d'une décision informatique doit toujours être entendue par un juge humain. »

(Table du Musée de la civilisation, Québec, 6 avril 2018, scénario Libération conditionnelle).

Dans la perspective préventive de l'IA utilisée à des fins policières est par ailleurs mentionnée une volonté de mettre en place un « cadre permettant de dépasser et d'éliminer les biais, discriminations et abus de pouvoir » (Table de la SAT, Montréal, 13 mars 2018, scénario Arrestation préventive) ainsi que le renforcement des lois sur le consentement afin que celui-ci soit véritablement éclairé. Il est également question de limiter l'accès des acteurs publics et privés aux données personnelles comme des « conversations privées sur les plateformes numériques » (Table de la bibliothèque Du Boisé, Montréal, 17 mars 2018, scénario Arrestation préventive) et de faire valoir un « droit à l'oubli, à la modification et à la correction des données ainsi qu'un droit à l'accès personnel aux données collectées » (Table de la bibliothèque Père Ambroise, Montréal, 3 mars 2018, scénario Arrestation préventive).

CODES DE DÉONTOLOGIE-D'ÉTHIQUE-DE CONDUITE

Les citoyens ont recommandé la mise en place d'une déclaration de principes, d'un code de déontologie, de conduite ou d'une procédure éthique au sein des entreprises, pour les différents ordres professionnels concernés ou tous les individus ayant accès à des algorithmes. Ces codes traiteraient de consentement, de confidentialité, de neutralité et du comment protéger la diversité humaine. Ils permettraient notamment de pallier la rapidité du développement des technologies d'IA et le caractère possiblement ingouvernable des entreprises qui la commercialisent.

> « Primauté de la déclaration de principes : Le vivre ensemble harmonieux. », c'est-à-dire qu'il faudrait toujours veiller à « une révision et une optimisation de l'algorithme en continu pour qu'ils soient toujours au service de l'humanité et de la diversité humaine »

(Table de la bibliothèque Père-Ambroise, Montréal, 3 mars 2018, scénario Arrestation préventive).

FORMATIONS

Les participants ont mis en avant la nécessité de développer des campagnes de sensibilisation afin de développer le sens critique des citoyens face à l'IA, à leur droit à la vie privée et au partage de leurs données. Cette éducation devrait porter aussi sur la littératie numérique et des compétences essentielles qu'il faut développer dès le primaire. Ces formations devraient assurer que les citoyens sont au courant des programmes et des types de données utilisés, de leur donner les connaissances et les outils nécessaires pour faire des choix informés, et mieux gérer les informations qu'ils partagent (ex. sous forme de campagne d'information, d'évènement public, de discussion).

Certaines tables ont également recommandé la mise en place d'une formation obligatoire pour tous les élèves du secondaire 1, 3 et 5 :

« L'apprentissage comprendra trois étapes :
1. Essence de l'IA
2. Fonctions et rôles de l'IA
3. Responsabilité éthique de l'IA. »

(Table de l'INM, Montréal, 18 février 2018, scénario Arrestation préventive).

Les citoyens ont également soulevé la nécessité de former les professionnels du secteur. Notamment, en recommandant que le conseil de la magistrature définisse les modalités de formation et adopte des règles pour former les juges aux nouvelles réalités technologiques, afin qu'ils comprennent comment fonctionnent l'IA, les enjeux éthiques reliés à l'IA et les conséquences d'une décision rendue par l'algorithme pour un individu et le professionnel.

MÉCANISMES PARTICIPATIFS

Les citoyens ont soulevé la nécessité de mener une grande consultation publique préalable à l'utilisation de l'IA en justice et à toutes mises en place d'encadrement. Le thème « Pour ou contre l'IA en justice » y serait mis au centre. Cette consultation viserait l'établissement de conditions particulières au développement de l'IA dans le secteur en amont de la mise en place des applications d'IA en justice. Cette consultation devrait être continue afin de suivre les nouveaux développements potentiels.

Les citoyens ont également proposé d'instaurer des mécanismes de prises de décision consensuels qui pourraient prendre la forme d'une coconstruction impliquant toutes les parties prenantes (ordres professionnels, associations, justiciables, ministère de la Justice, secteur industriel, etc.) lors de l'acquisition et du déploiement d'outils d'IA. L'emphase a été mise sur l'implication des usagers de l'IA dans ce secteur (ex. juges, avocats), qui doivent être impliqués dans la sélection du produit. En somme, une prise de décision consensuelle avec les parties prenantes lors de l'acquisition et du déploiement de l'outil a été jugée comme nécessaire.

PROGRAMMES DE RECHERCHE

Les citoyens recommandent également la mise en place de programmes ou centres de recherche universitaire, industrielles et multidisciplinaires portant sur les conséquences sociales, éthiques, économiques et politiques de l'IA sur notre société et la vie des individus. Selon les participants, il est indispensable de :

« S'assurer que la recherche génère des données probantes sur l'utilisation de l'IA en justice. »

(Table de la SAT, Montréal, 13 mars 2018, scénario Justice prédictive).

ACTEURS INSTITUTIONNELS ET AUTRES ACTEURS

En se demandant comment adapter les outils d'IA pour respecter les principes fondamentaux du système de justice, plusieurs participants ont évoqué la nécessité de créer un organisme indépendant certifiant les outils d'IA. Il ne s'agirait non pas de certifier la décision de l'outil, mais plutôt le processus décisionnel de l'algorithme. Cela permettait de s'assurer que les données soient exemptes de biais et que l'algorithme soit transparent et interprétable. La surveillance de la qualité de l'outil devrait se poursuivre après la certification, par un processus d'audit par exemple. Plusieurs tables ont évoqué que ces organismes indépendants pourraient prendre la forme d'entités hybrides (composées notamment d'acteurs publics-privés, d'ingénieurs, de professionnels du droit, de chercheurs en sciences sociales, de philosophes en éthique, etc.).

« Le but de cette entité serait le contrôle de l'IA. Elle identifierait les biais potentiels et serait coconstruite »

(Table de la SAT, Montréal, 13 mars 2018, scénario Justice prédictive).

Des participants ont aussi évoqué la nécessité de créer un groupe ou une instance indépendante qui devrait pouvoir agir comme un recours possible en cas de non-respect de certains principes liés aux droits fondamentaux ou de justice. À cet égard, le groupe pourrait être constitué de citoyens et de membres de la société civile. Dans la même optique, la création d'un ministère de l'éthique des données et de la protection numérique a également été proposée, notamment afin de préserver la diversité et le vivre ensemble harmonieux.

Enfin, pour assurer la liberté, la sécurité et la justice pour tous et toutes, des participants ont proposé de créer un « Centre de l'intelligence artificielle pour la sécurité civile » (CIASC). « Ce centre, composé de citoyens et de professionnelles », vise à contrôler « l'utilisation abusive de l'IA et met l'accent sur son rôle premier et but ultime qui est d'être un outil au service des citoyens. » (Table de l'INM, Montréal, 18 février 2018, scénario Arrestation préventive).

DISPOSITIFS D'ÉVALUATION DE L'IA

Les citoyens ont régulièrement mis de l'avant la nécessité de créer des normes et de produire des certifications (sur le processus de création et d'entraînement de l'algorithme) visant à protéger les droits et les libertés dans un contexte d'IA par des acteurs institutionnels. Il a aussi été question de mener des études pluridisciplinaires a priori et d'impacts a posteriori, de faire des tests, et de réviser et de maintenir l'algorithme à jour. Aussi, certains ont proposé de créer une certification « données claires et intention explicite » (Table de la bibliothèque Père-Ambroise, Montréal, 3 mars 2018, scénario Arrestation préventive). Il s'agirait d'une certification éthique sur la vulgarisation des données et sa finalité pour le corporatif et les ministères, notamment.

CADRES PROFESSIONNELS ET POLITIQUES INTERNES

Les participants ont exprimé leurs craintes que les entreprises commercialisant l'IA deviennent extrêmement habiles à échapper à tout contrôle. Leurs recommandations à ce sujet ont été de deux types. Premièrement, une procédure éthique devrait être implanter au sein des entreprises. Deuxièmement, les entreprises privées ou publiques devraient rédiger un rapport annuel obligatoire sur les incidents notoires liés à l'utilisation de l'IA, et ce, dans un souci de transparence.

MÉCANISMES ASSURANTIELS ET CONTRACTUELS

Les participants ont émis la nécessité que le secret industriel soit levé pour les acteurs judiciaires et les justiciables. Cela se ferait par le biais de contrats entre industriels et acteurs judiciaires qui indiqueraient la nécessité de rendre le code ouvert, examinable et vérifiable pour les acteurs du judiciaire et les justiciables.

> « Le code de l'IA devrait être ouvert et la décision devrait pouvoir être expliquée autant que possible »
>
> (Table de la SAT, Montréal, 13 mars 2018, scénario Justice prédictive).

5.4.

MONDE DU TRAVAIL

Tableau 7 : Pistes de solution ou grandes directions pour le secteur du monde du travail

	Nombre de pistes de solutions formulées
Formations	8
Acteurs institutionnels et autres acteurs	5
Dispositions légales	7
Incitatifs	3
Mécanismes participatifs	3
Codes de déontologie-d'éthique-de conduite	2
Outils numériques	1
Mécanismes assurantiels et contractuels	1
Dispositifs d'évaluation de l'IA	1
Programmes de recherche	1
Total	32

FORMATIONS

Concernant le milieu du travail, les citoyens ont recommandé la mise en place de formations pour tous afin de transmettre les connaissances nécessaires sur les enjeux actuels concernant le développement de l'IA. Ces formations devraient permettre de renforcer la littératie numérique et les compétences individuelles, et de garantir que les citoyens et les générations à venir soient sensibilisés, formés et prêts pour la transition numérique en cours.

Ces formations doivent tenir compte de la volatilité ou des incertitudes concernant le développement de l'IA dans le secteur du travail. Ceci pourrait se faire par la mise à jour des programmes scolaires, la mise en place de programmes de sensibilisation ou de soutien du gouvernement (ex. programmes de littératie numérique pour adultes) ou des formations continues pour les professionnels. Notamment, l'idée de la mise en place, par des agences gouvernementales, d'une *formation populaire* à l'IA et aux réalités numériques a été soulevée afin d'éviter l'exclusion sociale d'une partie de la population de son développement.

> « Programme de sensibilisation majeur des gouvernements du virage à l'IA et programme de soutien »
>
> (Table du Musée de la civilisation, Québec, 6 avril 2018, scénario Une restructuration responsable).

Afin d'éviter les enjeux liés à l'utilisation de l'IA dans le recrutement, les professionnels des ressources humaines devraient également suivre des formations rigoureuses sur les bases méthodologiques des algorithmes, la cueillette des données informatiques, le cadre légal qui accompagne cette collecte, les biais présents ou possibles dans l'analyse de l'IA. Un processus accéléré de mise à jour et de création de programmes professionnels devrait être planifié avec les cégeps, universités, ministères, ordres professionnels impactés par l'IA (ex. droit, santé).

ACTEURS INSTITUTIONNELS ET AUTRES ACTEURS

Les citoyens ont proposé la création de trois types d'acteurs institutionnels : une société d'État de l'IA au Québec, un comité interministériel conseillant le premier ministre et des comités de gouvernance dans toute entreprise ayant recours à des IA dans ses processus de recrutement.

La société d'État de l'IA au Québec, ou SNIAQ (Société nationale de l'intelligence artificielle au Québec), aurait pour mandats d'accompagner la transition numérique par son expertise en politiques publiques et par le soutien qu'elle pourrait apporter aux organismes privés et publics, et également en permettant un dialogue démocratique pour l'implantation de l'IA dans les services publics :

« Ses différents mandats sont :

> Assurer une expertise en IA, pour la conception des politiques publiques (travail, emploi, formations, aménagement du territoire, éducation, etc.) ;

> Organiser de manière démocratique l'expérimentation et l'implantation de l'IA dans la société et les services publics ;

> Soutenir les organisations privées et publiques face à la transition ;

> Soutenir et conseiller les ministères sur les programmes sociaux du Québec ;

> Aider le Québec dans les groupes de travail internationaux. »

(Table du Musée de la civilisation, Québec, 6 avril 2018, scénario Une restructuration responsable)

Le comité interministériel proposé serait un comité permanent et mixte, à l'interface des thèmes de l'économie, de l'emploi, de l'éducation et de la culture (inspiré de la Stratégie numérique). Il conseillerait directement le Premier Ministre. Ce comité permettrait au gouvernement d'avoir une expertise indépendante de consultants et qui ne dépende pas des entreprises privées ou d'organisations tierces.

Pour assurer de bonnes pratiques dans les entreprises en matière de recrutement assisté par IA, les comités de gouvernance proposés seraient quant à eux destinés à être formés dans chaque entreprise qui aurait recours à de l'IA dans ses processus de recrutement. Ces comités (un par entreprise) auraient pour mandat d'assurer le respect du code de déontologie des conseillers en ressources humaines (cf. « code de déontologie »). Il assurerait également la formation continue des recruteurs pour entretenir la vigilance envers des biais imprévisibles pouvant surgir à n'importe quel moment, et pour tenir compte du caractère évolutif de l'IA. Le comité de chaque entreprise serait multidisciplinaire, constitué d'experts dans le domaine de l'IA, d'experts en ressources humaines, et également de personnes travaillant hors du domaine des RH et de l'IA afin de permettre une diversité des avis et expériences, et une certaine indépendance. L'implantation d'un bureau de l'IA dans les entreprises a également été proposée afin de permettre aux employés de vérifier si l'utilisation d'une IA par un employeur est acceptable d'un point de vue légal.

DISPOSITIONS LÉGALES ET JURIDIQUES

Les dispositions légales et juridiques proposées par les participants tentent de répondre à deux principaux enjeux : la garantie d'un développement de l'IA centré sur l'humain avec la mise à jour de la charte des droits et libertés, et la protection (et révision) des données personnelles.

« Mettre à jour la charte des droits et libertés de la personne afin d'englober l'IA et la primauté de l'humain. »

(Table du Musée de la civilisation, Québec, 6 avril 2018, scénario L'IA comme passage obligé vers l'emploi)

En rapport au cadre légal, l'idée d'une redevabilité de l'entreprise a été défendue, notamment dans le contexte de la protection de la vie privée : dans le cas où un modèle prédictif est susceptible d'entrer en conflit avec le cadre légal actuel, l'entreprise responsable du modèle devrait communiquer l'information nécessaire pour juger de son impact. Toujours dans un souci de respect de la vie privée, la protection des données personnelles au travail pourrait être assurée par un règlement imposant d'informer l'usager du traitement de ses données et de l'informer des données détenues par l'entreprise et des individus en contact avec elles, à quelles fins, depuis et pour combien de temps. Ces informations devraient être accessibles et compréhensibles pour tout individu et pourraient être rassemblées dans un dossier numérique individuel (cf. « outils numériques »).

Par ailleurs, face au risque d'exclusion inhérent à la détention de données compromettantes, les participants invitent à permettre une forme de « réhabilitation numérique » de citoyens à qui certaines traces numériques porteraient préjudice. Pour encadrer cette sorte de droit à l'oubli, un cadre légal devrait être rédigé, notamment pour traiter des délais et des spécificités de cette réhabilitation numérique. Cela permettrait en même temps aux citoyens de choisir quelles informations les concernant sont disponibles, notamment sur les réseaux sociaux.

« Il faut respecter le cadre légal existant, à savoir les droits fondamentaux qui empêchent déjà les discriminations à l'embauche. On propose d'ajouter le droit à une réhabilitation numérique (ou droit à l'oubli) [afin que les gens ne soient pas injustement mis sur la touche pour des traces numériques consultées par des employeurs potentiels]. »

(Table de la SAT, Montréal, 13 mars 2018, scénario L'IA comme passage obligé vers l'emploi).

La création de lois de non-discrimination des algorithmes ou d'un revenu minimum garanti permettant de protéger les emplois perdus lors de la transition a également été discutée.

Les participants ont par ailleurs souligné la nécessité de l'adaptation du droit aux multiples enjeux accentués par l'IA, tout en entretenant une certaine souplesse dans le processus de révision des lois afin de pouvoir être réactif à l'évolution de l'IA et de ses effets. Des participants ont ainsi conseillé une « approche d'expérimentation » pour s'assurer de ne pas édicter des règles destinées à changer rapidement.

INCITATIFS

Les citoyens ont reconnu la nécessité de mettre en place différents incitatifs pour favoriser le développement responsable de l'IA dans le secteur du travail, en particulier en ce qui concerne l'enjeu de la transition numérique ainsi que la protection du bien-être des travailleurs. Ils ont d'abord soulevé la nécessité de repenser l'orientation des investissements publics concernant l'IA en société et d'exiger des investissements socialement responsables.

« Orientation des investissements vers une IA responsable pour le bien commun. »

(Table de la SAT, Montréal, 13 mars 2018, scénario Une restructuration responsable).

Provenant de l'État et des particuliers accompagnés par des conseillers publics en responsabilité sociale des entreprises, en synergie avec les fonds de travailleurs, ces investissements pourraient notamment prendre la forme de la mise en place d'un *Fonds pour la transformation numérique*. Les entreprises qui mettent en place des processus de transition pour leurs employés dont le travail se ferait remplacer par l'IA pourraient alors être subventionnées (ex. des formations avec un dispositif pour encourager ou exiger la fidélité de l'employé à l'entreprise une fois la formation terminée).

Dans ce même ordre d'idée, la création d'un fonds auquel entreprises et travailleurs cotisent a également été soulevée comme piste de solution qui pourrait conduire à la création d'une assurance numérique (voir « mécanisme assurantiel »). Celui-ci pourrait notamment permettre de répondre à l'enjeu de la précarisation par la mise en place d'un revenu minimum garanti.

Les citoyens ont également souligné la nécessité de revoir la structure des industries pour encourager l'inclusion des femmes (considérant l'intersectorialité), surtout si l'avenir du travail se situe dans ce secteur, afin de pallier le risque d'inégalités. Les citoyens ont ainsi proposé d'orienter l'attribution du financement selon un système de points qui encourage la diversité (une forme de *Quotas diversité* pour les entreprises encouragées par des politiques de renforcement plutôt que des sanctions).

Enfin, des citoyens ont encouragé le développement d'un programme de soutien à la création de nouveaux modèles d'entreprises de traitement des données, comme des coopératives qui auraient pour vocation de rompre l'isolement de travailleurs autonomes qui vont être de plus en plus nombreux.

De façon générale, dans un souci politique de partage des bénéfices de l'IA et afin de permettre une répartition équitable entre les groupes sociaux, les territoires et les différentes vulnérabilités, les participants invitent à développer une politique incitative de développement de l'IA qui lie responsabilité et subvention aux entreprises.

MÉCANISMES PARTICIPATIFS

Les participants ont proposé la création d'un espace permanent de concertation multisectoriel au sein du gouvernement pour répondre à l'enjeu de répartition des pouvoirs (lié au principe de démocratie) et aborder, entre autres, les questions de structuration des secteurs émergents.

Les citoyens ont également soulevé l'importance de la participation des usagers à la conception de l'interface des outils d'IA. Cette participation pourrait prendre la forme d'une expérimentation collective (design thinking) avec les différents partenaires. Elle leur permettrait de réviser le travail des programmeurs, notamment pour corriger des biais :

> « Permettre l'input usager dans l'apprentissage machine par une open AI (sur le modèle Wikipédia) pour corriger et réviser des biais par et pour la société. »

(Table du Musée de la civilisation, Québec, 6 avril 2018, scénario L'IA comme passage obligé vers l'emploi).

Le retour des usagers pourrait se faire auprès des autorités compétentes (ex. comités d'éthique, corporations) pour adapter le système.

CODES DE DÉONTOLOGIE-D'ÉTHIQUE-DE CONDUITE

Deux types de code de déontologie ont été proposés par les citoyens rassemblés autour du thème de la transformation du monde du travail : l'un à destination des conseillers en ressources humaines (CRHA) afin que les démarches de recrutement se fassent de façon non biaisée, l'autre à destination de toute profession utilisant des données personnelles à des fins commerciales — comme les publicitaires — afin que la protection de ces dernières soit mieux assurée.

Le premier, le code de déontologie des CRHA, répondrait à l'enjeu de « valoriser les diversités dans la construction des équipes » et découlerait des résultats d'un programme de recherche portant sur les biais de recrutement et mesurant l'impact de l'IA sur ceux-ci (cf. « Programme de recherche »).

Le deuxième code de déontologie cherche à répondre à l'enjeu de protection des données personnelles. Les participants invitent à « une réflexion sociétale sur l'utilisation des données personnelles » dans un contexte où ils jugent que les notions de « responsabilité » et de « bien commun » devraient faire l'objet d'un dialogue démocratique. Le code de déontologie en question découlerait de cette réflexion sociétale et démocratique et pourrait s'inspirer du règlement général européen sur la protection des données (RGPD).

« Au-delà du consentement individuel (ex. quand je visite un site internet), on doit avoir une réflexion sociétale sur l'utilisation des données et sur les enjeux de redistribution des richesses. »

OUTILS NUMÉRIQUES

Concernant le secteur du travail, un outil numérique a été proposé : la création d'un dossier numérique individuel. Celui-ci consisterait en un portail unique de nos données personnelles numériques qui s'accompagnerait d'une obligation de toute entreprise de déclarer les données qu'elle collecte. Ce type d'outil devrait être conçu de telle sorte que son fonctionnement soit transparent et compréhensible, notamment en ce qui concerne l'utilisation et la conservation des données personnelles.

MÉCANISMES ASSURANTIELS ET CONTRACTUELS

Pour accompagner la transition numérique et son impact sur le marché du travail, les citoyens ont proposé la création d'une assurance numérique sur l'IA pour permettre à chacun de se former et de s'adapter. Cette assurance serait financée par un fonds auquel les travailleurs et les entreprises cotisent (sur la base du même modèle que l'assurance parentale québécoise, adaptée à la réalité du travailleur). Il pourrait permettre d'obtenir de la formation, et ces formations seraient payées par les entreprises (dotées d'un dispositif encourageant, voire exigeant la fidélité de l'employé à l'employeur à la fin de la formation). L'assurance numérique pourrait également permettre de garantir un revenu minimum contre la précarisation des travailleurs dont l'emploi serait menacé.

DISPOSITIF D'ÉVALUATION DE L'IA

Des études d'impact ont été proposées afin de s'assurer que l'humain soit toujours considéré au centre de tout système d'IA. Celles-ci seraient réalisées par un organisme indépendant qui pourrait être financé par une taxe sur les données (sur le modèle de la taxe carbone).

« Dans l'analyse et la création de tout système, garantir et poursuivre une veille pour s'assurer de la primauté de l'humain par un organisme tiers indépendant (si nécessaire). Cet organisme serait financé par une data tax (comme carbon tax). »

(Table du Musée de la civilisation, Québec, 6 avril 2018, scénario L'IA comme passage obligé vers l'emploi)

PROGRAMMES DE RECHERCHE

Les participants ont recommandé le développement de programmes de recherche multidisciplinaires qui mesurent l'impact de l'IA sur les biais de recrutement. Ce programme de recherche informerait notamment la création d'un code de déontologie des conseillers en ressources humaines.

5.5.

SANTÉ

Tableau 8 : Pistes de solution ou grandes directions pour le secteur de la santé

	Nombre de pistes de solutions formulées
Dispositions légales et juridiques	11
Acteurs institutionnels et autres acteurs	9
Dispositifs d'évaluation de l'IA	8
Formations	6
Codes de déontologie-d'éthique-de conduite	4
Programmes de recherche	4
Cadres professionnels et politiques internes	2
Mécanismes participatifs	1
Outils numériques	1
Total	46

DISPOSITIONS LÉGALES ET JURIDIQUES

En ce qui concerne les lois et les règlements, plusieurs recommandations ont été émises à des niveaux distincts ; notamment en matière de vie privée, de transparence, de collecte de données ou encore concernant l'universalité des soins de santé.

Pour plusieurs citoyens, s'il est primordial de s'appuyer sur les lois et règlements existants en matière du droit de chacun au contrôle de ses informations personnelles, il faut aussi penser à la manière donc ils pourraient être redéfinis pour prendre en compte les innovations technologiques liées à l'IA. La garantie de la vie privée a été un élément important lors des discussions. À cet égard, les citoyens expriment la nécessité de garantir la confidentialité des informations personnelles.

« Des lois devraient être mises en place afin de garantir la propriété privée des données personnelles (p. ex., une loi donnant accès aux données collectées aux personnes concernées) »

(Table de l'INM, Montréal, 18 février 2018, scénario Jumeaux Numériques).

Certaines tables ont aussi indiqué la nécessité de mettre en place des lois et règlements favorisant la transparence et la clarté des objectifs visant la collecte, l'utilisation et l'accès des données

biologiques (et tout autre renseignement personnel sur la santé). Ces informations devraient être accessibles dans un format clair et compréhensible pour l'utilisateur. Des participants ont soulevé la nécessité d'émettre des directives concernant les organismes gouvernementaux collectant des données biologiques et personnelles sur la santé en ce qui a trait à l'intelligibilité, la qualité et la pertinence des informations transmises.

En ce qui concerne la collecte des données, il est nécessaire d'encadrer la provenance ou les sources qui servent à l'algorithme pour garantir qu'il n'y ait pas de biais pour les citoyens. Pour maintenir un système de santé qui soit juste, les participants recommandent aussi d'émettre des lois et règlements en ce qui concerne les objectifs du système de santé, à savoir le respect du principe de l'universalité des soins :

> « Inscrire sous la loi d'accès universel aux soins tous les développements de l'IA en santé au même titre que les solutions alternatives »

(Table de la bibliothèque Mordecaï-Richler, Montréal, 10 mars 2018, scénario Vigilo).

Dans le contexte canadien, la proposition d'établir une loi spécifiant si (et comment) la couverture des soins publics offerte par la RAMQ pouvait s'appliquer aux innovations technologiques liées à l'IA dans le domaine de la santé a été soulevée.

Enfin, une réglementation relevant du Collège des médecins a également été proposée afin que l'humain puisse toujours prévaloir sur l'IA.

> « L'utilisation de robots ne doit pas se faire sans la supervision d'une autorité (humaine) institutionnelle soumise à une déontologie. »

(Table de la bibliothèque Mordecai-Richler, Montréal, 10 mars 2018, scénario Vigilo).

ACTEURS INSTITUTIONNELS ET AUTRES ACTEURS

En ce qui concerne la santé, les citoyens ont recommandé la mise en place de plusieurs comités et acteurs institutionnels. Par exemple, il pourrait s'agir de comités consultatifs dont la mission serait de définir les « valeurs » que devrait prendre en considération l'IA dans son traitement de l'information. Les citoyens ont émis l'idée de créer un organisme indépendant capable de statuer sur les bénéfices et les risques sur la vie privée et qui, en parallèle, se pencherait autant sur les questions de santé que sur les questions éthiques reliées aux enjeux de l'IA. Il faudrait aussi, selon les participants, former des comités pour réviser les erreurs commises par des dispositifs d'IA qui permettraient d'améliorer les algorithmes. Il pourrait notamment s'agir d'une obligation pour le système de santé d'évaluer périodiquement la validité de ses algorithmes, de rendre leur fonctionnement et cette évaluation publique avec une clause de « déclaration de toutes modifications » (Table de la bibliothèque Père-Ambroise, Montréal, 3 mars 2018, scénario *Jumeaux numériques*).

Pour certains participants, il est nécessaire de désigner quelqu'un de légalement responsable afin qu'une personne soit imputable en cas de fautes commises. Dans cette lignée, il faudrait aussi avoir un endroit pour faire appel à une décision émise par un algorithme. Mettre en place un ombudsman indépendant dont la fonction serait de régler les litiges entre patients et médecins a également été proposé.

Pour d'autres citoyens, la nomination d'un commissaire à la vie et au bien-être qui « statue sur les objectifs de la santé tout en défendant les citoyens et la population en général, et notamment le droit de ne pas savoir » est nécessaire (Table du Musée de la civilisation, Québec, 6 avril 2018, scénario *Jumeaux numériques*). La création d'une instance pour mettre en place un cadre de gouvernance humaniste et indépendant du développement de l'IA dans l'application des soins de santé a également été proposée. Enfin, les citoyens ont recommandé la mise en place d'un centre d'anonymisation des données de santé géré par le gouvernement qui aurait pour objectif de protéger les citoyens du détournement de leurs données personnelles par des entreprises privées.

DISPOSITIFS D'ÉVALUATION DE L'IA

Les citoyens ont recommandé la mise en place d'une certification éthique des IA en santé, soit le développement de certifications (ou labels) pour les algorithmes et robots, sur la base de données issues de projets de recherche (ex. recherche participative, sur le contexte qui influence le développement de l'IA) pour déterminer les critères de cette certification et ses différents niveaux. Ces critères devraient inclure la transparence, la sécurité et la pertinence de l'outil. Par exemple, ces certifications auraient pour but d'uniformiser l'accès aux processus décisionnels des algorithmes ou de valider les outils des robots de santé. Ces certifications devraient être attribuées par le gouvernement ou des organismes indépendants et multipartites, afin de protéger le bien-être des patients et les intérêts publics, et viseraient principalement les entreprises privées qui développent les IA en santé.

« Certification en amont des robots de santé et de leur trousse d'outils (notamment, pour protéger les intérêts du public) »

(Table de la bibliothèque Mordecai-Richler, Montréal, 10 mars 2018, scénario Vigilo).

FORMATIONS

Concernant le secteur de la santé, les participants ont reconnu la nécessité de mettre en place des mesures éducatives et de sensibilisation pour toutes les parties prenantes du développement de l'IA en santé, incluant les professionnels de santé et le public. Les formations professionnelles, qui pourraient prendre la forme de formations continues (ex. sur la base de la création d'un guide de bonnes pratiques) devraient porter, entre autres, sur la relation médecin-patient-IA, avec des études de cas et statistiques à jour. Ces formations auraient pour objectif d'assurer non seulement une utilisation optimale et consciente des algorithmes, mais aussi une communication adéquate et juste de l'information à fournir aux patients afin d'éviter les mésinterprétations.

Concernant la formation du public, les participants ont recommandé qu'une conscientisation se fasse dès le début de l'éducation des jeunes générations (à l'école), afin de favoriser le développement d'un sens critique vis-à-vis des technologies d'IA. L'idée d'un cours d'autodéfense intellectuelle a été soulevée pour développer ce sens critique et éduquer les utilisateurs aux nouvelles pratiques par le biais de vulgarisation.

« Dès le niveau primaire, débuter de conscientiser les jeunes générations et favoriser le développement d'un sens critique. S'assurer de la justesse des informations transmises au public et celles qui méritent d'être communiquées au citoyen / patient »

(Table du Musée de la civilisation, Québec, 06 avril 2018, scénario Jumeaux numériques).

CODES DE DÉONTOLOGIE-D'ÉTHIQUE-DE CONDUITE

Les citoyens ont également recommandé l'adoption de codes d'éthique ou de déontologie, qu'ils soient destinés à toute entreprise qui crée de l'IA pour la santé ou, plus globalement, à l'attention des usagers et professionnels de santé sur le territoire canadien. Ces codes doivent contenir des normes en ce qui concerne la sécurité, la transparence et la responsabilité des médecins ou des développeurs. Ces codes devraient permettre d'assurer que chaque citoyen soit accompagné par un médecin pour toute décision médicale. Certains citoyens ont mentionné qu'il fallait inclure la définition de la responsabilité humaine face à l'IA dans les codes de déontologie déjà existants. Par exemple, a été proposée la mise en place d'un serment d'Hippocrate 2.0 qui implique un professionnel de santé dans toute recommandation de santé afin de garantir une personnalisation des soins et un accompagnement humain. Celui-ci pourrait impliquer la mise en place de « garde-fous virtuels » qui empêcheraient l'algorithme de dévier et de biaiser le diagnostic.

« Le code de déontologie et la responsabilité du médecin doivent toujours prévaloir sur l'IA. Cette dernière n'est qu'un outil d'aide. »

PROGRAMMES DE RECHERCHE

Les citoyens ont recommandé la mise en place, le financement et l'encouragement de différents programmes de recherche interdisciplinaires en matière d'IA appliquée à la santé. Les participants sont tous d'avis que l'on doive mettre de l'avant la recherche en IA, mais également d'autres domaines qui étudient les effets de l'IA sur la société comme les sciences sociales, la philosophie ou la bioéthique. Ces recherches devraient par exemple permettre d'identifier le partage des responsabilités en amont et en aval des différentes parties prenantes, de mesurer l'impact de l'IA sur leur autonomie ou d'alimenter les programmes de formation et d'éducation des intervenants comme des citoyens.

« Élaborer des programmes de recherche pour évaluer à quel point la situation socioéconomique d'un individu a un impact sur sa santé et sur les éventuels diagnostics d'une IA »

(Table de l'INM, Montréal, 18 février 2018, scénario Jumeaux numériques).

CADRES PROFESSIONNELS ET POLITIQUES INTERNES

En réponse au risque d'atteinte à la vie privée, les citoyens recommandent que le système de santé soit responsable de documenter et rendre transparent pour le patient l'accès à ses données par des tierces parties (« qui » et « quand »).

Également, une des recommandations discutées par les citoyens visait une procédure à suivre pour le diagnostic (suivant l'idée d'un diagnostic double humain-machine). Cette procédure inciterait le médecin à poser son diagnostic avant de connaître celui de l'algorithme, ce qui permettrait de protéger l'expertise et l'indépendance du médecin et que l'algorithme reste un outil complémentaire d'aide qui peut informer et aider le médecin dans sa prise de décision. Cet algorithme ne tiendrait pas seulement compte des données purement médicales (ex. indicateurs biologiques) du patient, mais d'autres types de données (ex. mode de vie, alimentation).

MÉCANISMES PARTICIPATIFS

Les citoyens ont mis en évidence la nécessité de tenir un débat et une consultation publique sur la sécurité des données avant le dépôt d'un ou de plusieurs projets de loi. Ces débats devraient inclure le public, des experts et d'autres parties prenantes actuellement déjà impliquées (ex. éthiciens).

« Il faut sortir du contexte du simple citoyen face à son ordinateur qui est confronté à une politique de confidentialité. »

OUTILS NUMÉRIQUES

La création d'un *formulaire de consentement électronique* adapté à la réalité numérique a été proposée. Il serait convivial, numérique et interactif, et toujours accompagné d'une personne ressource.

5.6.

VILLE INTELLIGENTE ET OBJETS CONNECTÉS

Tableau 9 : Pistes de solution ou grandes directions pour le secteur de la ville intelligente et des objets connectés

	Nombre de pistes de solutions formulées
Dispositions légales et juridiques	14
Acteurs institutionnels et autres acteurs	10
Formations	10
Dispositifs d'évaluation de l'IA	5
Mécanismes participatifs	5
Outils numériques	3
Codes de déontologie-d'éthique-de conduite	2
Incitatifs	1
Programmes de recherche	1
Total	51

DISPOSITIONS LÉGALES ET JURIDIQUES

Les participants aux tables abordant le thème de la ville intelligente et des objets connectés ont proposé la mise en place de plusieurs dispositions légales et juridiques. Ces pistes de solutions ont comme objectif d'assurer la protection des données personnelles et du consentement des usagers ainsi que la loyauté de la technologie. Par exemple, concernant le contrôle des objets connectés, les citoyens ont proposé un règlement qui autoriserait la déconnexion en tout temps. Également, face à différents risques (dont celui d'invasion de la vie privée) les participants invitent à réfléchir sur la possibilité d'envisager une disposition légale sur la loyauté des objets connectés, qui garantit que les mesures prises et les recommandations faites sont dans l'intérêt du consommateur et non celle de la compagnie :

> « Loi définissant la notion de loyauté et autres considérations éthiques (discrimination) »

(Table de la SAT, Montréal, 13 mars 2018, scénario Jouet intelligent).

Concernant l'utilisation de la technologie par des mineurs, les citoyens ont recommandé la détermination légale d'un âge de « maturité numérique » :

> « Il faudrait penser à un âge de raison numérique. À la maturité numérique. »

Cette mesure fait écho à la proposition d'encourager le développement d'une « citoyenneté numérique » qui permettrait de responsabiliser le citoyen face aux changements dictés par les nouvelles technologies. Celle-ci permettrait notamment de définir les responsabilités et former les citoyens sur leurs droits et devoirs en matière d'accessibilité de l'IA.

L'idée de la mise en place d'un moratoire a également été soulevée par les citoyens. Celui-ci pourrait être d'une durée d'un ou deux ans et devrait permettre d'encadrer d'un point de vue légal l'utilisation de l'intelligence artificielle dans le cadre du transport public :

> « Avant l'implantation, on devrait encadrer la situation. Un moratoire est nécessaire maintenant sur la mise en chantier tant que nous n'aurons pas une technologie qui soit responsable. »

Afin de tenir compte des enjeux d'équité, les participants proposent la mise en place d'un *programme social d'aide à la mobilité* qui permettrait d'éliminer les barrières de l'accès à l'IA pour certaines catégories de personnes au statut précaire. Dans la même optique, les citoyens ont proposé la mise en place d'un droit à la mobilité visant le transport pour tous. Une réforme des lois concernant le transport, le code de la route et la sécurité routière a ainsi été proposée. Les citoyens considèrent également comme nécessaire de revoir les lois en matière d'urbanisme, par exemple en mettant en place une réglementation où il conviendrait d'inciter l'aménagement mixte avec la prise en compte d'une certaine diversité.

La mise en place d'une règlementation qui permet de sécuriser les données personnelles et le partage d'information a également été recommandée. Cette règlementation permettrait de protéger l'anonymat, la propriété des données, d'assurer le respect de la vie privée ou porterait sur l'interdiction de la captation des données hors du service prévu. Ces lois devraient également permettre une meilleure transparence en ce qui concerne le traitement des données personnelles par le secteur privé.

> « Élargir la loi sur le consentement afin de garantir que l'individu reste propriétaire de ses données. »

(Table du Musée de la civilisation, Québec, 6 avril 2018, scénario Réfrigérateur connecté).

Ces lois, selon les citoyens, devraient être intégrées à la constitution canadienne. Afin de préserver les choix des utilisateurs sur les paramètres de transport, les citoyens ont proposé la mise en place de lois fédérales tout en préservant une règlementation adaptable au niveau local.

ACTEURS INSTITUTIONNELS ET AUTRES ACTEURS

Les participants aux tables abordant le thème de la ville intelligente et objets connectés ont proposé plusieurs idées pour la nomination d'acteurs institutionnels, qu'il s'agisse de sociétés indépendantes ou de comités consultatifs. L'idéal démocratique de comités ou assemblées permettant la participation citoyenne a été rappelé à plusieurs reprises.

Pour le contrôle des objets connectés, deux modèles ont ainsi été proposés, comprenant un mécanisme forçant l'autorégulation des acteurs privés :

- Sur le modèle de la Régie du logement du Québec, une *Régie des objets connectés* permettrait de fixer les prix des objets connectés (tels que les réfrigérateurs) et proposerait une aide sociale pour en favoriser l'acquisition. Elle émettrait également des certificats de propriété à l'achat d'un objet connecté pour établir que les données générées par cet objet appartiennent à l'usager. Celui-ci choisirait alors de donner ou non son accord pour que ces données soient communiquées à la compagnie commercialisant cet objet ainsi qu'à son assureur, et ce, sans pénalité.

> Une autorité indépendante sur la gestion des données pourrait permettre aux citoyens un recours collectif quand il y a des usages abusifs. Elle pourrait également gérer une plateforme numérique pour que les utilisateurs puissent parler librement et publiquement des avantages et inconvénients des dispositifs d'IA et avoir ainsi un effet sur l'image de marque des acteurs privés commercialisant ces objets. Ceux-ci seraient ainsi amenés à s'autoréguler par la pression des utilisateurs exercée sur leur image.

Pour répondre à un enjeu d'équité et ainsi assurer un partage équitable de l'IA, un *défenseur des droits* pourrait être joignable au « 1-800 discrimination des objets connectés » (Table de l'INM, Montréal, 18 février 2018, scénario *Réfrigérateur connecté*). Il pourrait faire partie d'un « comité multipartite gérant démocratiquement les incidents, les injustices et autres enjeux » (Table de la bibliothèque Mordecai-Richler, Montréal, 10 mars 2018, scénario *Voiture autonome*). Par ailleurs, un commissaire aux comptes indépendant pourrait également être mandaté pour faire un audit comptable afin d'assurer un partage équitable des bénéfices de l'IA.

Concernant la régulation des véhicules autonomes, la création de la SAIAQ (Société de l'assurance de l'intelligence artificielle du Québec) apporterait des modifications aux lois de sécurité routière pour les adapter à la conduite autonome. Elle comprendrait également une dimension d'assurance automobile 2.0 qui proposerait de nouvelles formes de contrats pour ce type de conduite (Table de la bibliothèque du Boisé, Montréal, 17 mars, scénario *Voitures autonomes*).

Pour une organisation efficace des réseaux de la ville intelligente et optimiser le système urbain géré par l'IA, les participants ont proposé un organisme hybride, le MIAOU (Mobilité, intelligence artificielle et optimisation urbaine) financé par le gouvernement du Québec (Table de la SAT, Montréal, 13 mars 2018, scénario *Voiture autonome*). Ce centre aurait pour mission de gérer et optimiser le développement de l'ingénierie responsable de l'IA et rassembler les connaissances pour permettre d'aider à la prise de décision concernant la rédaction des règlements et des lois suite aux projets pilotes.

Les participants ont également envisagé la formation de différents groupes de personnes, tels qu'un *ministère du développement technologique* qui gèrerait les directives pour le département du territoire intelligent qui encadrerait, lui, les transformations urbaines en lien avec l'IA et la ville durable ; ou encore une *commission de défense du droit à la mobilité pour les voitures autonomes* afin de garantir la protection du droit à la mobilité (cf. Dispositions légales et juridiques).

FORMATIONS

Les participants ont recommandé la mise en place de formations pour les citoyens sur les nouvelles technologies liées à la ville intelligente, afin de mieux comprendre le fonctionnement de l'IA et les nouvelles normes qui l'accompagnent. Cette éducation pourrait prendre la forme de vulgarisation, de formations continues ou de sensibilisation. Elle porterait, par exemple, sur le fonctionnement et l'utilisation de l'IA ou le civisme et la ville numérique.

Les participants ont recommandé une formation à la vigilance collective pour une utilisation responsable de l'IA. Ces formations devraient viser à démocratiser l'information sur l'IA afin de responsabiliser les individus sur ses règles d'usage, favoriser les choix éclairés et permettre aux utilisateurs de participer à la prise de décision.

« Cours d'éducation à la littératie des données à plusieurs niveaux de la scolarité pour que les citoyens aient les armes et réflexes nécessaires pour faire des choix éclairés »

(Table de la SAT, Montréal, 13 mars 2018, scénario Jouet intelligent)

L'éducation à l'IA dans le secteur de la ville doit se faire à tous les niveaux et dans différents lieux (ex. dans le cadre de bibliothèques, de coop, de *fab lab,* de l'école ou d'organisme à but non lucratif). Elle pourrait prendre la forme de cours pratiques dans les écoles pour pouvoir apprendre à gérer les différents objets connectés ou de programmes d'éducation à la littératie des données.

DISPOSITIFS D'ÉVALUATION DE L'IA

Les citoyens ont reconnu la nécessité de mettre en place des dispositifs d'évaluation des coûts, des « effets de bord » et des impacts de politiques spécifiques à l'IA. Ils ont envisagé la possibilité de mettre en place des normes (ex. label éthique) qui protègent le consommateur, ramènent l'être humain au centre des décisions et favorisent l'inclusion. Par exemple, les citoyens ont proposé la mise en place d'une norme similaire à la certification ISO qui permettrait une reconnaissance des entreprises qui offrent des services numériques à valeur ajoutée pour les citoyens. L'objectif d'une telle norme serait de garantir que l'utilisateur ait le choix et reste en contrôle des services qu'il utilise afin de garantir que ces derniers ne soient pas intrusifs.

Également, la création d'une certification qui permettrait d'assurer une collaboration entre humains et machine a ainsi été proposée. Celle-ci devrait garantir la sureté, la sécurité, l'exploitabilité, la transparence, la loyauté et/ou la confiance des utilisateurs :

« Certification mesurant et garantissant le niveau de loyauté et autres considérations éthiques de mon objet connecté »

(Table de la SAT, Montréal, 13 mars 2018, scénario Jouet intelligent).

MÉCANISMES PARTICIPATIFS

Les citoyens ont recommandé la mise en place d'assemblées publiques qui pourraient prendre la forme d'un *forum hybride démocratique* d'expression des citoyens, pour l'évaluation des projets et le diagnostic des besoins des usagers, visant à déterminer l'aménagement de l'espace public selon les besoins de chacun et les valeurs de la société. Également, les citoyens ont proposé la mise en place d'un *système de recours collectifs* pour les usages abusifs, qui permettrait une agilité dynamique pour s'adapter au progrès technique.

D'autres propositions qui impliquent la participation active des citoyens ont été présentées, comme la mise en place de sondages, d'un planning participatif (évaluation du plan d'urbanisme pendant la période de transition), de systèmes, voire de code de déontologie *open source* (pour trouver des solutions aux enjeux collectifs et qui vise à l'amélioration du bien-être de la collectivité). Les citoyens ont soulevé la nécessité de revoir la distribution des compétences entre provincial, municipal et arrondissement.

OUTILS NUMÉRIQUES

Les participants ont proposé d'intégrer dans la conception des objets connectés un développement permettant de comprendre facilement et de visualiser les données qu'un ou des objets génèrent (à qui/quand/où ils les communiquent et pourquoi), afin de garantir et faciliter la personnalisation de leurs réglages. Il s'agit ici de garantir une conception pluridisciplinaire des objets connectés qui intègre les aspects émotionnels et psychologiques du rapport à l'alimentation, ou autre, de l'individu dans le processus de design (cf. scénario *Réfrigérateur connecté*) ou de proposer des choix de trajets par rapport à des critères personnels (cf. scénario *Voiture autonome*).

CODES DE DÉONTOLOGIE-D'ÉTHIQUE-DE CONDUITE

Les citoyens ont également recommandé la mise en place d'un code de déontologie pour les informaticiens et concepteurs d'IA, qui pourrait être mis en place et suivi par un organisme indépendant et qui aurait pour objectif de statuer sur le devoir de transparence et la traçabilité, l'inclusion et la prise en compte des risques dans l'optique de protéger l'intérêt général. Ce code prendrait ainsi la forme d'un permis de responsabilité pour défendre le bien commun.

INCITATIFS

Les citoyens ont reconnu la nécessité de mettre en place des incitatifs pour encourager les compagnies à divulguer leurs sources, leurs biais, les algorithmes qu'elles utilisent et ainsi assurer la transparence des recommandations et des actions des objets connectés (p. ex., par des déductions d'impôts ou des appels d'offres). Ces incitatifs (qu'ils soient individuels ou collectifs) pourraient également encourager l'utilisation d'autres moyens de transport (cf. scénario Voiture Autonome). Par exemple, ces incitatifs pourraient prendre la forme d'un système de Points mobilité pour les individus encourageant le partage des transports, en particulier ceux qui fonctionnent à l'aide d'énergie verte ou à basse production de gaz à effet de serre.

PROGRAMMES DE RECHERCHE

Les participants ont soulevé la nécessité de mener des études pour comprendre les implications de l'utilisation de l'IA et garantir un développement harmonieux de la société à différents niveaux ainsi que de réfléchir à la conservation du patrimoine humain.

« Mener des études pour comprendre les implications de l'utilisation de l'IA et garantir un développement harmonieux de la société (psychologie, culture, enjeux sociaux, égalité, éducation »

(Table de la SAT, Montréal, 13 mars 2018, scénario Jouet intelligent).

La mise en place de projets pilotes capables d'encourager le déplacement collectif dans la ville et de prendre en considération les enjeux d'équité sociale tout en permettant l'élimination des barrières architecturales a également été proposée.

MÉCANISMES ASSURANTIELS

Bien que ces recommandations n'aient pas été formulées sur les affiches, les participants ont proposé la mise en place d'une assurance numérique afin d'assurer l'intégrité et de protéger la propriété des données personnelles, qu'il s'agisse de leurs usages dans le cadre des voitures autonomes ou des objets connectés. Par ailleurs, la création de nouveaux types de contrats pour les usages liés à la conduite automobile a été proposée pour assurer une bonne gestion de l'IA dans le cadre des transports individuels.

6. CONCLUSION

À l'issue de cette délibération ayant réuni plusieurs centaines de citoyens, usagers et experts, de nombreux enjeux et pistes de solution ont été identifiés. L'objectif de la démarche étant d'entendre les citoyens sur le développement responsable de l'IA, les discussions se sont organisées autour de scénarios mettant en scène plusieurs risques, et soulevant différents enjeux éthiques préalablement identifiés – faisant écho aux principes de la Déclaration de Montréal. Cette observation devrait permettre de tempérer un certain scepticisme à l'égard du développement de l'IA qui peut ressortir de ces résultats, sans pour autant le négliger. Ces résultats nous donnent une certaine idée de l'acceptabilité sociale de l'IA et de son développement.

La grande diversité des propositions implique de pousser plus loin l'analyse en vue de recommandations pour des politiques publiques. L'ensemble des résultats présentés ici nous permet en effet de soulever différentes problématiques qui méritent une analyse approfondie en vue de formuler nos recommandations. Se pencher sur ces problématiques est apparu comme essentiel afin de se prononcer sur un encadrement responsable du développement de l'IA. Elles sont abordées à travers 4 chantiers qui nous ont paru prioritaires :

1. Les enjeux de la gouvernance des SIA
2. Le développement de la littératie numérique de l'ensemble des citoyens
3. L'inclusion de la diversité dans le développement des SIA
4. La promotion d'une soutenabilité forte du développement des SIA

Le développement de l'IA soulève ainsi de nombreuses questions de société. Si ces questions ne sont pas toutes forcément propres à l'IA, les transformations portées par son développement dans différentes sphères sociales amènent à se questionner, en tant que citoyens, sur la société à construire. Au cœur des tensions entre espoirs et craintes, c'est la relation entre humain et technologie qu'il est essentiel de souligner. Si une revendication semble faire consensus, c'est bien celle de conserver une place centrale de l'humain dans un monde de plus en plus artificiellement intelligent.

< >
Déclaration de Montréal
IA responsable_
</ >

PARTIE 4

LA COCONSTRUCTION DE L'AUTOMNE 2018 : LES ACTIVITÉS CLÉS

TABLE DES MATIÈRES

RÉDACTION

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

CHRISTOPHE ABRASSART, professeur à l'École de design, Université de Montréal

CAMILLE VÉZY, candidate au doctorat en communication, Université de Montréal

COLLABORATION

LOUBNA MEKKI BERRADA, doctorante en neuropsychologie clinique, Université de Montréal

VINCENT MAI, doctorant en robotique, Université de Montréal

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

1. INTRODUCTION

Les principales activités de coconstruction de la Déclaration de Montréal se sont déroulées du 3 novembre 2017 au 31 avril 2018. Toutefois, l'équipe de la Déclaration a poursuivi ses activités lors de l'automne 2018 en organisant trois activités clés afin de mobiliser les connaissances d'un plus grand nombre d'acteurs sur différents thèmes essentiels lorsqu'il est question du développement responsable de l'intelligence artificielle (IA). Ainsi, une journée de coconstruction à Paris a été organisée selon le modèle des coconstructions de l'hiver 2018. Afin de mobiliser les connaissances de parties prenantes du secteur culturel, un groupe de discussion sur les enjeux liés à l'avènement de l'IA dans les domaines relatifs à l'art et la culture a également discuté de ces questions. Enfin, afin de faire le pont entre les délibérations citoyennes de l'hiver et la relève en recherche, une activité de simulation de rédaction de brèves politiques avec des étudiants des cycles supérieurs a été réalisée en partenariat avec le Comité intersectoriel étudiant (CIÉ) des Fonds de recherche du Québec (FRQ).

Ces différentes activités sont venues alimenter l'analyse et la rédaction des recommandations formulées en vue de politiques publiques (cf. rapport *Les chantiers prioritaires et recommandations pour un développement responsable de l'IA*). Les sections suivantes récapitulent les enjeux identifiés et les principales pistes de solutions formulées par les personnes ayant participé aux activités. Certaines reprennent les mécanismes proposés lors de la coconstruction de l'hiver et viennent d'autant plus appuyer la nécessité de leur mise en place, tandis que d'autres ont un caractère inédit.

2. JOURNÉE DE COCONSTRUCTION HORS QUÉBEC (PARIS, FRANCE)

Cette section présente les résultats de la journée de coconstruction réalisée à Paris le 9 octobre 2018, organisée en partenariat avec l'Ambassade du Canada à Paris, le Centre culturel canadien et la Maison des étudiants canadiens. Lors de cette journée, 26 personnes aux profils variés ont été mobilisées afin de se pencher sur les enjeux du développement responsables de l'IA. Les participants ont été regroupés autour de trois tables de coconstruction, chacune abordant un thème clé du développement de l'IA : la démocratie, l'environnement et le monde du travail.

Figure 1 : Profils des participants à la journée de coconstruction de Paris.

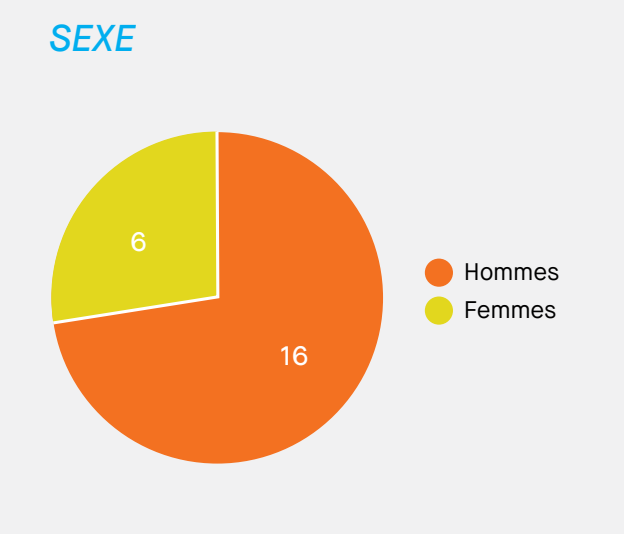


Figure 2 : Profils des participants à la journée de coconstruction de Paris.

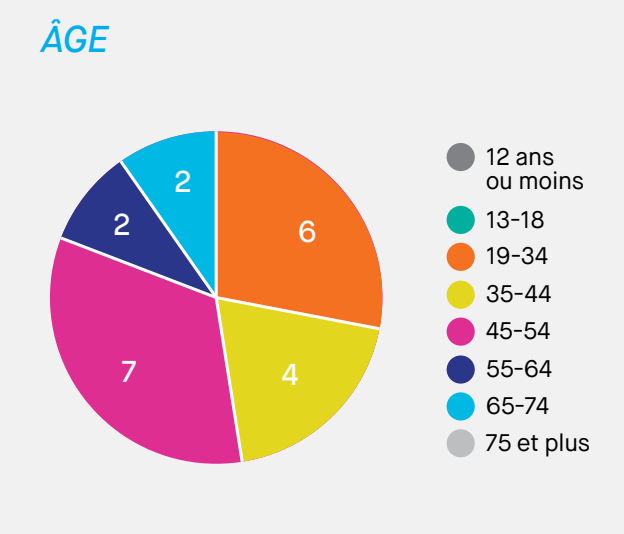


Figure 3 : Profils des participants à la journée de coconstruction de Paris.

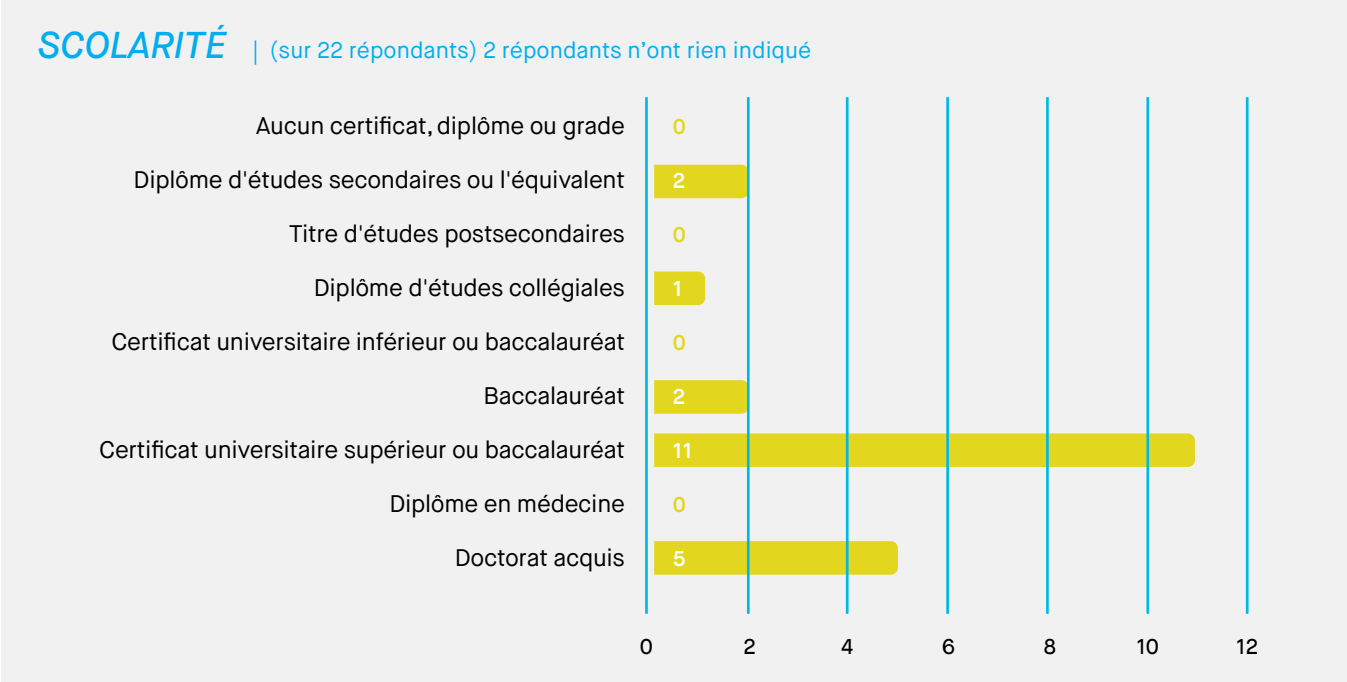
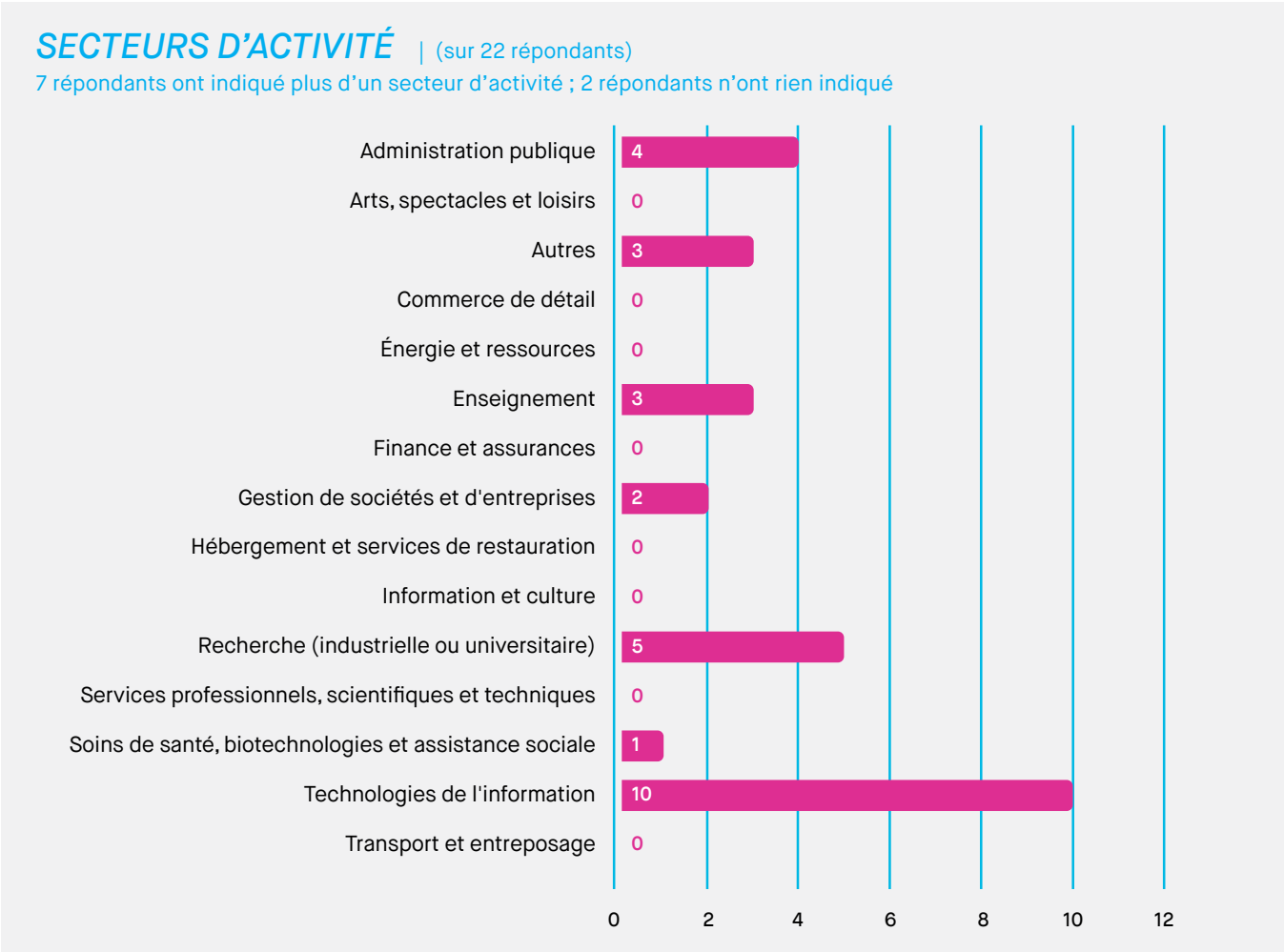


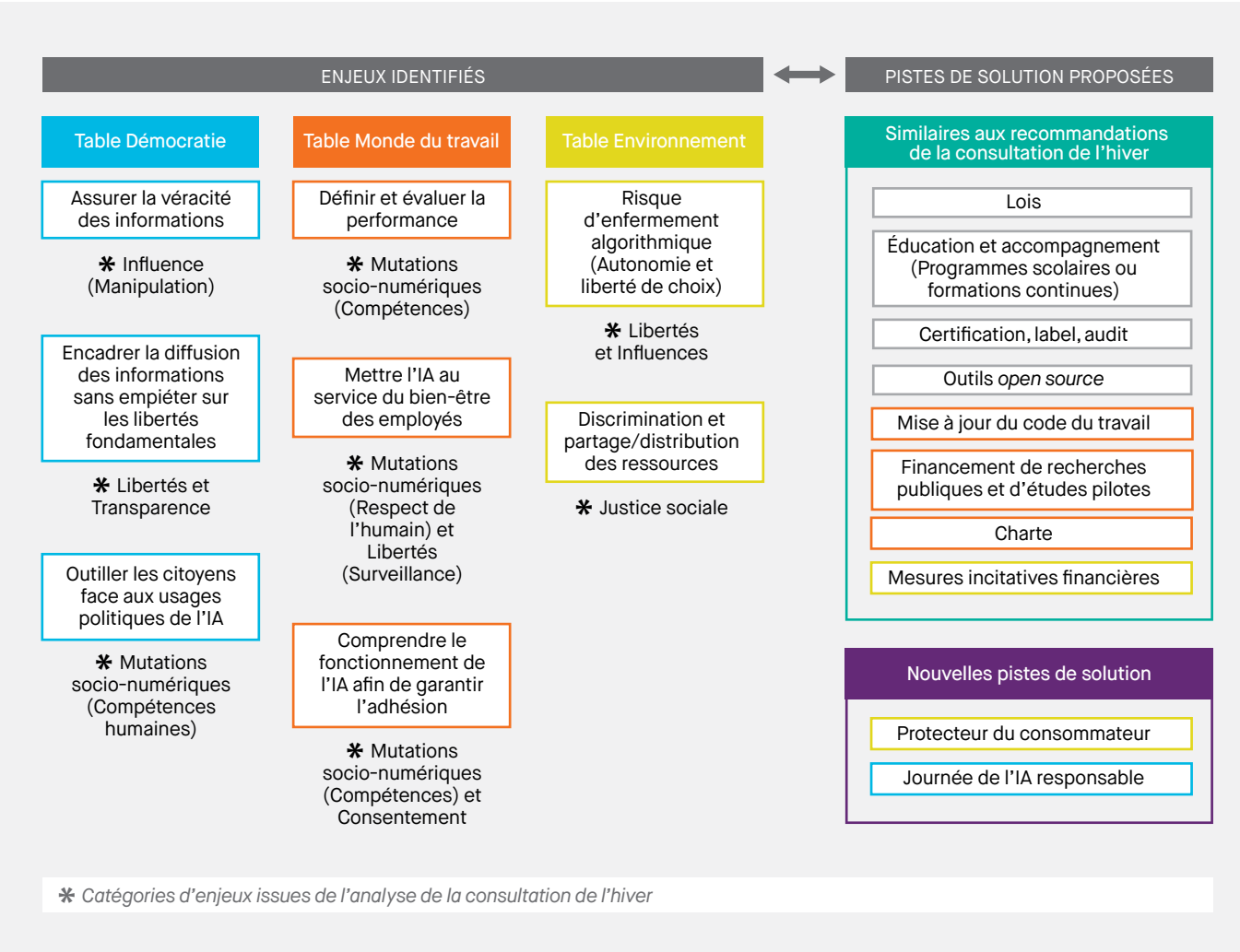
Figure 4 : Profils des participants à la journée de coconstruction de Paris.



Les discussions se sont organisées autour de trois moments délibératifs (identifications des enjeux, formulation de recommandations et écriture de la « Une » d'un journal), suivant le modèle de coconstruction développé pour les activités de l'hiver. Trois scénarios déclencheurs ont ici été utilisés (cf. Annexe 1) mettant en scène l'utilisation de l'IA dans le cadre de l'incitation aux comportements écologiques, de la gestion de personnel dans l'entreprise, et de l'utilisation de l'IA pour la création de fausses nouvelles dans le cadre d'une campagne électorale. Ces trois scénarios nous ont permis d'explorer les enjeux relatifs à la démocratie, à l'environnement et au monde du travail sous un angle nouveau, n'ayant jamais été utilisé sous cette forme précédemment.

Les sections suivantes retracent le parcours de chacune des discussions ayant eu lieu autour des trois tables[1]. Ces discussions ont mis en évidence l'apparition d'enjeux déjà identifiés lors de la coconstruction de l'hiver, mais selon les spécificités des contextes donnés. Également, différentes pistes de solution ou mécanismes d'encadrement pour un développement responsable de l'IA ont été proposés. Certains sont similaires à ceux formulés lors des discussions de l'hiver (ex. Lois et Formations) et d'autres sont nouveaux (ex. Journée de l'IA responsable).

Schéma 1 : Enjeux soulevés et pistes de solution proposées lors de la journée de coconstruction de Paris



[1] Les citations sont tirées de notes rédigées par des participants.

2.1.

ABORDER LES ENJEUX DÉMOCRATIQUES À TRAVERS LES FAUSSES NOUVELLES

Résumé du scénario de départ

Fausse nouvelle dans la campagne électorale

À deux semaines des élections présidentielles, une réunion de crise a lieu au sein de l'Agence sur l'intégrité de l'information (AII) mise en place dans le cadre de la Loi contre la manipulation de l'information. Une vidéo diffusant des propos compromettants du président sortant sur des travailleurs immigrés est devenue virale. Le porte-parole de la présidence a fait savoir qu'il s'agissait d'un faux créé par une agence étrangère qui tentait d'interférer dans les élections en tirant parti des algorithmes GAN (*generative adversarial networks*). Bien que la diffusion de la vidéo soit interdite, celle-ci continue à circuler via différents sites étrangers.
À un mois du premier tour des élections présidentielles, l'AII doit présenter un plan pour enrayer les effets dévastateurs de cette fausse information et rétablir les conditions d'une campagne électorale saine.

L'objectif de ce scénario était d'ouvrir une discussion sur les enjeux éthiques liés à la manipulation de l'information pouvant nuire aux démocraties par leur diffusion virale. Et ce, en particulier quand des techniques d'intelligence artificielle permettent d'imiter des personnes et modifier des contenus en conservant un très haut degré de réalisme rendant particulièrement difficile la détection du faux.

Le parcours délibératif présenté est issu d'une journée complète de discussion regroupant sept chercheurs, experts et étudiants travaillant en éthique, en développement organisationnel, en machine learning, en web social et en sciences politiques. Partant de ce scénario en 2022, la discussion a conduit à la formulation d'une « Une » du Journal de l'IA responsable du 9 octobre 2022 : « Première journée citoyenne de l'IA responsable ».

Premier moment délibératif : Formulation d'enjeux éthiques en 2022

DÉMOCRATIE

Les participants sont d'avis que l'IA en elle-même n'est pas la cause d'atteintes à la démocratie : ces problèmes existent déjà, ils sont cependant nuancés ou amplifiés par les possibilités permises par l'IA. Il faut donc se préparer et s'adapter à cette nouvelle réalité. « La santé de la démocratie » a alors été avancée comme un enjeu particulièrement important dans un contexte où les « choix basés sur une information erronée » sont problématiques. L'encadrement par le droit et la formation de l'esprit critique ont été abordés comme deux pistes de réflexion pour se prémunir contre les effets de la manipulation de l'information et ainsi veiller à la santé de la démocratie.

ENCADREMENT PAR LE DROIT

Les participants se sont demandé comment les pratiques liées à la manipulation de l'information pourraient être encadrées :

« Le droit peut-il encadrer l'entièreté des pratiques de l'IA (manipulation de l'information) ? »
Un participant

Dans cette perspective, la question du « comment faire justice » quand la réputation est entachée par la propagation de fausses nouvelles a été mentionnée plusieurs fois. Cela se heurte cependant à la difficulté d'encadrer la manipulation de l'information sans entraver la liberté d'expression et les autres libertés fondamentales.

CONNAISSANCE ET ESPRIT CRITIQUE

« [L'] importance de la corrélation démocratie/éducation (notamment l'éducation à la pensée critique) » a été soulignée, marquant un appel au développement de l'esprit critique pour la « participation éclairée à la vie publique ». Un participant précise que plutôt que de se prémunir contre la propagande elle-même, le développement de l'esprit critique permet de se prémunir contre les effets de la propagande (endoctrinement, modifications de choix et comportements, polarisation vers les extrêmes, etc.) En effet, empêcher toute propagande, c'est-à-dire toute action exercée sur l'opinion, risquerait de mener à la censure ou d'entraver la liberté d'expression. Il faudrait donc plutôt miser sur l'éducation des citoyens pour développer la « pensée critique » ainsi que la « littératie médiatique et statistique ».

L'enjeu de démocratisation de l'accès à l'information et à la connaissance a par la suite été soulevé. Plusieurs conditions sont ici nécessaires : que la neutralité d'internet soit assurée afin de garantir le libre accès à toute information, et que tous aient accès à des outils technologiques permettant de se renseigner et de s'exprimer. Chacun, sans discrimination, devrait ainsi pouvoir être sensibilisé au fonctionnement de l'IA et à ses enjeux.

AUTHENTIFICATION ET CAPACITÉ DE L'IA POUR IDENTIFIER LE FAUX

En apprenant certaines techniques d'IA, plusieurs pourraient par exemple devenir capables de développer des outils de détection et de correction de fausses nouvelles. Cela dépend cependant de la possibilité et de la nécessité d'utiliser l'IA pour identifier le faux du vrai dans un contexte où il est par exemple impossible de distinguer à l'œil nu la différence entre une vidéo créée par IA et une vidéo originale. Les participants se sont ainsi interrogés sur les « mécanismes d'authentification des nouvelles/informations » ou encore sur la « capacité de la technologie à comprendre les sarcasmes ». La « traçabilité des sources sur tous les outils de diffusion (ex : WhatsApp) » a été par exemple proposée comme piste de mécanisme d'authentification. Il a cependant été précisé qu'une IA qui censurerait automatiquement les publications considérées comme fausses, malveillantes ou non fiables aurait un effet négatif sur la liberté d'expression.

RESPONSABILITÉ

Face à l'impression de « perte de contrôle de l'information » dans un contexte de propagation très rapide, voire virale des fausses nouvelles, l'enjeu de la responsabilité a été abordé à plusieurs étapes : création, diffusion, partage, lecture des informations. Les participants se sont demandé à quel moment la véracité des nouvelles devrait être évaluée (avant sa création, avant sa publication, avant son partage, pendant sa lecture, etc.), par quelle entité (organisation internationale, État, journalistes, lecteurs, plateforme de diffusion qui permet la publication et/ou le partage tel que WhatsApp ou Facebook), et de quelle manière (développement de l'esprit critique du lecteur, création d'un label indiquant le niveau de véracité de l'information et de fiabilité de la source).

La création et la diffusion de fausses nouvelles font ainsi appel à une combinaison des rôles de plusieurs acteurs devant assumer des responsabilités quant au fait de créer des informations, les diffuser, juger de leur crédibilité, vérifier leur véracité et la fiabilité des sources qu'ils soient journalistes, lecteurs ou autres. Cette entité devrait évaluer la véracité des nouvelles, la plateforme de diffusion d'information ainsi que tout individu qui crée ou diffuse des informations à titre personnel.

Les participants ont ainsi discuté de la « déontologie des médias (captation et diffusion d'informations) » et de la « crédibilité du journalisme ». Un nouveau rôle des journalistes et médias pourrait être de juger l'information publiée et partagée en vérifiant et déclarant si celle-ci provient ou non d'une source fiable.

Tableau 1 : Démocratie, Premier moment délibératif : formulation d'enjeux éthiques en 2022

Enjeux éthiques 2022	1	2	3
Description	Assurer la véracité des informations	Encadrer la diffusion des informations sans empiéter sur les libertés fondamentales	Outiller les citoyens face aux usages politiques de l'IA
Principes associés	Démocratie, responsabilité	Démocratie, autonomie	Démocratie, connaissance, autonomie

À la suite de cette discussion, les participants ont formulé ces trois enjeux prioritaires en proposant des pistes d'encadrement afin de veiller à la santé de la démocratie face aux effets de la propagande via la création et la diffusion automatisées de fausses informations :

1. Assurer la véracité et la fiabilité des informations, notamment pour préserver la santé des échanges démocratiques.
2. Encadrer la diffusion des informations sans empiéter sur les libertés fondamentales, dont la liberté d'expression, en particulier par le développement de normes journalistiques et technologiques en matière de diffusion de l'information.
3. Outiller les citoyens quant aux usages politiques de l'IA, pour que ceux-ci puissent se renseigner et développer leur propre opinion librement.

Deuxième moment délibératif : Propositions d'encadrement de l'IA

Pour répondre à ces enjeux, l'équipe a poursuivi ses discussions pour formuler ensemble cinq propositions d'encadrement de l'IA :

Tableau 2 : Démocratie, Deuxième moment délibératif : propositions d'encadrement de l'IA pour 2018-2020

Propositions d'encadrement	1	2	3	4	5
Description	Création d'une autorité de certification et de normes journalistiques via un label (signal rouge, jaune, vert)	Création d'outils *open source* pour détecter/ distinguer le faux du vrai. (ex : application sur téléphone mobile)	Projet de loi garantissant la neutralité du web.	Mise en place de programmes scolaires + formation continue (par exemple MOOC) pour le développement de l'esprit critique	Mise en place d'une journée de l'IA citoyenne responsable
Enjeu associé	Assurer la véracité des informations	Encadrer la diffusion des informations sans empiéter sur les libertés fondamentales		Outiller les citoyens face aux usages politiques de l'IA	

Dans la lignée des réflexions précédentes, les discussions menant à ces propositions ont porté sur les dimensions techniques, politiques et éducatives du problème de la manipulation de l'information. Les participants ont ainsi proposé un ensemble de mesures visant à encadrer la vérification des nouvelles (tables 1 et 2) et à éduquer les citoyens pour participer à la vie démocratique de façon libre et éclairée (tables 3, 4 et 5). Cela passe par la normalisation des pratiques de création et diffusion de l'information (certification) et l'aide à la détection du faux (outil), l'accès égal de tous à l'information via un réseau internet neutre (projet de loi), l'éducation pour tous, à tous les âges de la vie pour le développement de l'esprit critique (programmes scolaires et formations continues) et la sensibilisation aux usages politiques de l'IA (journée de l'IA responsable).

La création d'une autorité de certification établissant des normes journalistiques via un système d'indicateur de fiabilité des informations a émergé de la réflexion autour de la responsabilité des journalistes pour établir la fiabilité des sources et la crédibilité des informations. Les participants ont été d'avis que cet organisme de certification devait être indépendant du gouvernement.

Plutôt qu'un encadrement juridique, une approche par les normes professionnelles et l'établissement de lignes directrices à la création, la vérification et la diffusion des informations a été préférée. Un participant a souligné qu'il était plus raisonnable d'adopter une approche minimaliste dans la conception de ce système indicatif de la fiabilité d'une information. En effet, il faudrait éviter qu'un tel système influence davantage les processus électoraux et autres situations de compétition où les nouvelles jouent un rôle crucial. Ce serait le cas par exemple si le système indicatif de fiabilité pouvait donner avantage à un parti ou être utilisé comme une stratégie contre l'adversaire dans une perspective de manipulation ou de lobbyisme. En même temps, en optant pour une approche minimaliste, on pourra mieux éviter des limitations excessives risquant de porter atteinte aux libertés fondamentales.

En optant pour cette solution préventive, les participants n'ont pas proposé de solutions réactives en cas de diffusion d'une fausse nouvelle. Ils ont toutefois mentionné la nécessité dans ce cas d'interrompre le plus rapidement possible la propagation et de contrecarrer les fausses nouvelles par leur version vérifiée.

Les pratiques journalistiques de vérification de nouvelles instaurées par cette autorité de certification pourraient avoir recours à un outil *open source* dont la technologie permettrait de distinguer le faux même quand celui-ci est indétectable à l'œil nu, en particulier quand ce faux est créé par des techniques d'IA. Ces pratiques devraient aussi permettre de vérifier que certains renseignements véridiques ne soient pas validés comme étant des fausses nouvelles.

Soulignant également la part de responsabilité importante des plateformes de partage d'information comme les médias sociaux (Facebook, Twitter, Youtube, Snapshat, etc.) et les messageries instantanées (ex. : WhatsApp), les participants se sont interrogés sur la nécessité d'obliger ces plateformes d'instaurer un outil faisant appel à l'IA pour identifier les fausses informations. Des participants ne semblaient pas faire confiance aux entreprises privées pour instaurer un système d'identification et de blocage de fausses nouvelles qui soit juste, étant donné que pour l'instant ce type de plateforme laisse en place des vidéos contenant de faux renseignements qui influent sur l'opinion des différents publics. Un système indiquant le niveau de fiabilité d'une nouvelle serait cependant intéressant à voir apparaître sur ces plateformes qui doivent être absolument responsabilisées par rapport au rôle qu'elles jouent dans la propagation des informations manipulées qui affectent les démocraties.

La formation et la sensibilisation des citoyens font l'objet d'un deuxième ensemble de mesures visant le développement de l'esprit critique des citoyens, en particulier par la littératie médiatique qui les outillerait pour naviguer de façon libre et éclairée dans l'univers de l'information. Cela peut passer à la fois par les curriculums scolaires et par les espaces publics ouverts à tous, via des universités populaires dans des bibliothèques ou des cafés par exemple, ou bien des campagnes créatives de sensibilisation qui s'immisceraient dans la vie quotidienne pour entretenir la vigilance quant aux pratiques de manipulation de l'information.

Troisième moment délibératif : Écriture de la « Une » du Journal en 2022

Ces propositions ont ensuite été mises en récit, à la « Une » du Journal de l'IA responsable du 9 octobre 2022 formulée ainsi :

Première journée citoyenne de l'IA responsable

Dans le cadre de la première journée internationale de l'IA responsable, plusieurs événements ont été organisés simultanément au Canada et en France, notamment afin d'informer et outiller les citoyens face aux nouvelles possibilités offertes par l'IA et à l'importance de la pensée critique. À cet effet, notre journal est très fier de vous informer qu'il a reçu son accréditation de l'Ordre des journaux IA responsable.

En projetant la lumière sur la « Journée citoyenne de l'IA responsable », les participants soulèvent ainsi l'importance de la sensibilisation et l'appropriation des enjeux de l'IA par l'ensemble des citoyens pour pouvoir participer à la vie démocratique. La mention de l'obtention de l'accréditation de la publication à titre de « journal IA responsable » traduit également un souci de renforcement du rôle des médias et du journalisme à titre d'acteurs importants de la santé démocratique.

2.2. ABORDER LES ENJEUX LIÉS À L'ENVIRONNEMENT

Résumé du scénario 2025 de départ

Les températures continuent à battre des records de chaleur dans le monde entier. Pour faire face à cette crise du changement climatique, des villes proposent à leurs habitants un système de permis carbone individuel fortement incitatif, le système EcoFit, qui est connecté à leur compte bancaire et aux différentes applications d'achat en ligne : dans ces villes, le prix des biens et services est affiché en euros et en carbone, et chaque citoyen doit viser 4 tonnes d'émission de carbone par an pour l'ensemble de sa consommation. Cette cote leur donne un accès à de multiples services écoresponsables en transports, éducation, formation et culture. Au cours de l'année 2025, Ive et Charles réussissent à ajuster progressivement leur consommation sur cette cible, et même à faire mieux. Et comme ils ont moins dépensé, ils ont réalisé une épargne inattendue. Ils imaginent alors un projet de séjour à Cuba pour Noël et commencent à consulter les sites de voyages. Un message leur parvient alors sur leur téléphone : « Attention à l'effet rebond : dépenser vos économies dans un voyage en avion annulerait tous vos efforts ! Pensez à voyager local ! »

L'objectif de ce scénario était d'ouvrir une discussion sur les possibilités autant que sur les enjeux éthiques d'une gestion prédictive par SIA des effets rebonds environnementaux sur les marchés de biens de consommation et d'équipements. Les effets rebonds peuvent s'expliquer de la manière suivante. Alors que l'efficacité énergétique des équipements

s'améliore, alors que l'empreinte environnementale des biens de consommation diminue grâce à l'écoconception, plutôt que de verrouiller ces acquis, on consomme proportionnellement plus d'équipements, de biens et de services : par exemple la taille des écrans augmente, le taux d'équipement des ménages augmente, on parcourt plus de km en automobile, on voyage en avion, etc. Il en résulte une augmentation des émissions de GES et une pression accrue sur les ressources et la biodiversité. Avec ces effets rebonds, il n'y a donc pas de découplage entre le développement économique d'une part, et sa matérialité et son empreinte écologique d'autre part.

Le dispositif algorithmique imaginé ici se situe dans une approche d'utilisation des SIA pour la planète (« AI for Earth »). La possibilité d'une gestion prédictive et personnalisée des effets rebonds, par apprentissage supervisé sur des historiques de consommation (ex. des données de transactions bancaires), associée à des dispositifs incitatifs (nudges) a ainsi été mise en scène dans ce scénario exploratoire.

Le parcours délibératif présenté est issu d'une table tenue en 3 h avec huit citoyens intéressés par les nouvelles technologies et les enjeux d'environnement et de développement durable. Partant de ce scénario en 2025, la discussion a débouché sur la formulation d'une initiative présentée en « Une » du Journal de l'IA responsable du 9 octobre 2020 : « Grand succès pour ConsoM'IA : 1 millon d'abonnés en une semaine. Règlement général d'ouverture des data pour l'environnement ».

Quel a été le parcours délibératif de ce groupe pour mener à cette proposition originale ? Quels ont été ses moments marquants ? Comment se sont enrichies les idées à chaque étape ? Nous présentons, en les commentant, certains moments significatifs du parcours suivi par cette équipe.

Premier moment délibératif : formulation d'enjeux éthiques en 2025

De nombreuses interrogations rédigées sur des *Post-it* ont été formulées par les participants en relation avec différents principes de la Déclaration de Montréal :

LE PRINCIPE DE VIE PRIVÉE

« Pourra-t-on retracer tout l'historique de consommation de cette famille ? », « La gestion des bases de données sera-t-elle assez fiable pour protéger les données personnelles et donner confiance aux usagers ? », « Pourrait-il y avoir un droit à l'effacement ? ».

LE PRINCIPE D'AUTONOMIE ET DE LIBERTÉ DE CHOIX

Ce dispositif conduit-il à un nouveau « pouvoir prescriptif ? », « Permet-il de conserver une autonomie dans la décision et le libre arbitre ? », « Comment avoir un jugement critique sur ces recommandations personnalisées ? », « Assiste-t-on au contrôle de la vie courante par une machine ? », « Y a-t-il un risque d'enfermement algorithmique, de bulles algorithmiques ? », « Comment un tel système peut-il prendre en compte le contexte singulier d'une décision d'achat (ex. une situation d'urgence) ? ».

LE PRINCIPE DE RESPONSABILITÉ

Ce dispositif doit aider à « renforcer la responsabilité environnementale au quotidien », « à vivre une éthique personnelle comme ConsoActrice », mais avec le recours à des outils d'IA, « y a-t-il un risque d'externalisation de la responsabilité individuelle ? »

LE PRINCIPE DE JUSTICE ET ÉQUITÉ

Ce dispositif, en demandant aux entreprises d'évaluer l'empreinte carbone de leurs produits et services préalablement à leur introduction sur le marché, permettra-t-il « d'assurer une libre

concurrence ? », « Risque-t-il de renforcer le pouvoir des grandes entreprises et de créer une barrière à l'entrée discriminante pour les PME, avec le coût de ces bilans environnementaux ? », « Comment le commerce équitable, qui apporte d'autres dimensions éthiques, sera-t-il évalué ? », « Certains producteurs seront-ils privilégiés ? ». « Les riches pourront se permettre de consommer plus et d'acheter des quotas de carbone pour compenser leurs émissions ! C'est une inégalité sociale ! » Par ailleurs, si « tout doit passer par les données et le marché », « les initiatives de réduction des GES hors marché (ex. projet d'habitants de quartier en mobilité active, en agriculture urbaine) seront-elles rendues invisibles et donc discriminées ? ». Enfin, « les styles de vie sont différents dans le monde, les régimes alimentaires sont variés (ex. végane, religieux), risque-t-on d'en favoriser certains et d'en discriminer d'autres ? », « de créer des discriminations culturelles ? ».

LE PRINCIPE DE DÉMOCRATIE ET GOUVERNANCE

« Qui va réguler ce système ? Est-ce les Nations Unies ? Est-ce les pays riches ? Comment surveiller les abus ? », « Devrait-on avoir une autorité régulatrice sur les bilans de carbone ? », « Si on fait une économie de CO2, pourrait-on transmettre nos économies de carbone à nos proches ? », « Les recommandations à prioriser devraient-elles être discutées de manière démocratique ? ».

Plusieurs discussions approfondies ont ensuite eu lieu, les participants rebondissant sur les premières idées pour en générer d'autres. Puis, après près de 45 minutes de discussion, les participants ont sélectionné, à l'aide de pastilles colorées, des regroupements d'enjeux éthiques pour 2025 qui leur semblaient prioritaires. Deux principes de la Déclaration de Montréal sont ainsi ressortis : autonomie, conjugué à la liberté de choix, et justice, associé par les participants au principe d'équité.

Tableau 3 : Environnement, Premier moment délibératif : formulation d'enjeux éthiques en 2025

Enjeux éthiques 2025	1	2
Description	Risque d'enfermement algorithmique face à ce nouveau pouvoir de prescription et à la configuration des espaces de choix pour chacun. Comment garder une autonomie individuelle et collective ? Comment valoriser une initiative de réduction des émissions de carbone hors du système ?	Les plus riches pourraient être favorisés par la compensation carbone : quelle limite leur assigner ? Inversement, ceux qui consomment peu pourraient-ils redistribuer leur épargne de carbone ? Quelles relations entre pays du Nord et du Sud ? Y a-t-il des risques de discrimination culturelle ?
Principes associés	Autonomie et liberté de choix	Justice et équité

Cette sélection d'enjeux prioritaires par l'équipe conduit à développer plus précisément deux enjeux éthiques des SIA. Le premier est lié au principe d'autonomie, avec la possibilité d'actions visant la réduction des gaz à effet de serre hors marché (ex. une initiative citoyenne sur la mobilité quotidienne). Le second a trait au principe de justice, avec la possibilité de compensation carbone pour les plus riches, ou de partage des émissions pour les citoyens ayant une consommation plus restreinte.

Deuxième moment délibératif : propositions d'encadrement de l'IA pour 2018-2020

Pour répondre à ces enjeux, l'équipe a poursuivi ses discussions en essayant de réfléchir ensemble aux quatre principes associés. Plusieurs propositions d'encadrement de l'IA ont été formulées par les participants. Nous en présentons ici trois, qui permettent de suivre le cheminement des idées jusqu'à la formulation de la « Une » du Journal.

Tableau 4 : Environnement, Deuxième moment délibératif : propositions d'encadrement pour 2018-2020

Propositions d'encadrement en 2018-2020	**1**	**2**	**3**
Description	• Établir un code déontologique pour les concepteurs, programmeurs et gestionnaires du système (ex. pour assurer l'égalité des prescriptions)	• Le protecteur du consommateur, autorité administrative indépendante, auditée par l'Assemblée nationale. • Audits du système, de la diversité de choix et des recommandations, et publication de rapports transparents. • Accompagnement des citoyens pour leur autonomie.	• Assurer un soutien financier significatif pour aider les personnes les plus modestes à s'adapter. • Avoir la possibilité de dépasser la cible annuelle, mais avec un coût marginal croissant sur la tonne de carbone supplémentaire.
Catégories d'instruments	Loi et règlement Code de déontologie	Acteur institutionnel	Mesures incitatives et d'accompagnement

Ces propositions, qui dénotent une véritable créativité institutionnelle (au-delà des exemples d'instruments très généraux donnés dans le livret du participant) se situent dans la lignée des enjeux identifiés à l'étape précédente. La proposition d'un protecteur du consommateur (assurant une évaluation régulière du système par des audits, rendant publiquement des comptes et organisant un accompagnement des citoyens), présente ainsi un enrichissement des idées formulées à l'étape précédente. C'est à partir de cette proposition que les participants vont bâtir leur « Une » du Journal lors de l'étape suivante.

Troisième moment délibératif : écriture de la « Une » du Journal en 2020

Ces mesures ont ensuite été mises en récit dans l'affiche. La « Une » du Journal de l'IA responsable du 9 octobre 2020 formulée par l'équipe était la suivante :

Grand succès pour ConsomIA ! Un million d'abonnés en une semaine

Règlement général d'ouverture des *datas* pour l'environnement

Suite à la loi RGODE (Règlement général d'ouverture des *datas* pour l'environnement), obligeant à rendre publiques les données personnelles, l'association CONSOM'IA a mené une enquête d'envergure sur la liberté de choix des usagers d'ECOFIT et constaté de nombreuses limitations et bulles algorithmiques. La première recommandation du rapport de CONSOM'IA est de développer des moyens de contre-expertise, et la formation des usagers pour assurer un vrai pluralisme et la participation de tous à la réduction des gaz à effet de serre.

Ainsi, si l'utilisation de l'IA présente un certain potentiel pour gérer les enjeux environnementaux associés aux comportements de consommation, cette perspective soulève également de nombreux enjeux éthiques qui doivent être encadrés convenablement.

2.3.

ABORDER LES ENJEUX DE LA TRANSITION NUMÉRIQUE DANS LE MONDE DU TRAVAIL

Résumé du scénario de départ

Forage des données (*Data mining*) RH pour optimiser l'ambiance au travail

Pierre-André a enfin décroché un emploi dans un bon bureau d'avocats. Après trois semaines de travail, il rencontre Marco aux ressources humaines pour une séance de mentorat personnalisée. Il lui explique que la firme utilise désormais AmbIA+, une IA d'analyse conversationnelle qui étudie les attitudes des salariés et aide à maintenir une ambiance de travail apaisante et productive (tous les courriels, appels téléphoniques et prises de parole en réunion d'équipe sont analysés). AmbIA+ fournit une assistance individualisée, elle conseille et entraine, mais il n'y a pas de sanction. Tous les échanges que Pierre-André a eus jusqu'à présent au bureau se sont bien passés, sauf les 15 et 16 octobre derniers. « Vous avez à plusieurs reprises interrompu vos collègues en réunion pour répéter les mêmes idées, ce qui a créé de la tension chez eux. Apparemment, l'algorithme a aussi détecté des périodes d'inactivité sur le réseau de plusieurs heures, sans aucun échange avec vos collègues. Ce n'est pas grave en soi, mais c'est mieux de maintenir le contact avec l'équipe. Est-ce que vous vous souvenez de la raison de cette inactivité sur le réseau ? ». Pierre-André n'est plus seulement inquiet, il est embarrassé et s'interroge sur la pertinence de ces questions.

L'objectif de ce scénario était d'entamer une discussion sur les enjeux éthiques liés à l'utilisation de l'IA dans le cadre de la surveillance et de la gestion du personnel au sein des entreprises. Notamment, le système AmbIA+ imaginé ici est utilisé afin d'optimiser la performance et de contrôler l'ambiance au travail par le biais de techniques de *data mining*.

Le parcours délibératif présenté est issu d'une journée complète de discussion regroupant 10 ingénieurs, concepteurs d'IA, gestionnaires en stratégie numérique, chercheurs, étudiants et professeurs. Partant de ce scénario en 2025, la discussion a mené à la formulation d'une « Une » du Journal de l'IA responsable du 9 octobre 2025 : « Le premier employé renvoyé à cause de l'IA ».

Premier moment délibératif : la formulation d'enjeux éthiques et sociaux en 2025

Lors de cette première partie, les participants ont identifié cinq catégories d'enjeux en lien avec le développement de l'IA dans le monde du travail.

AUTONOMIE

D'abord, l'enjeu du respect de l'autonomie (notamment en ce qui a trait à la capacité d'agir des employés) a été souligné. Les participants ont dénoncé une certaine manipulation de « la manière dont les gens ressentent les choses », et une culture organisationnelle « forcée ». Les participants ont considéré que la sauvegarde d'informations, telles que les comportements et interactions entre employés dans le but de cultiver une culture organisationnelle, est problématique. En fonctionnant ainsi, l'entreprise réalise une certaine normalisation des employés, qui pourrait conduire à de fortes tensions (voire « un totalitarisme » si tout venait à être mesuré afin d'assurer un contrôle de l'entreprise sur les individus), par le biais de recommandations insistantes de l'IA. Le respect de l'autonomie est lié au respect d'un certain « libre arbitre » des employés et du respect de leurs émotions qui, ici, ne sont pas suffisamment considérées selon les citoyens. Est-il bien de conserver toutes ces données (la moindre action étant observée) et de les utiliser dans ce but ? Les participants ont ainsi souligné un enjeu lié à la « surveillance », qui pourrait alors limiter le champ d'action et la liberté de parole des employés (très lié à l'enjeu du respect de la vie privée).

« L'IA t'observe »
Un participant

Les citoyens ont mis de l'avant la nécessité de favoriser l'autonomie en permettant à chacun de travailler à sa façon pour le meilleur de l'entreprise. L'employé devrait être capable de garder le contrôle sur ses données et l'employeur devrait l'informer de la collecte et de l'utilisation de ces dernières avec précision, et chacun devrait être libre de pouvoir se « déconnecter », notamment afin de préserver la frontière entre ce qui relève du professionnel et du privé.

VIE PRIVÉE

La place de cette frontière entre vie privée et professionnelle au sein de l'entreprise a été débattue par les participants et considérée comme floue. D'un côté, pour certains, le système d'IA en jeu porterait atteinte à la vie privée des employés, en écoutant leurs conversations :

« Qu'en est-il des discussions personnelles ? »
Un participant

De l'autre, certains participants ont mentionné qu'habituellement, tout ce qui se rattache à la vie personnelle ne devrait pas se faire ni par le courriel ni par le téléphone du travail (et donc ne serait pas dans la mire de l'analyse du SIA présenté dans le scénario). Ces participants ont soulevé la question suivante : Y a-t-il une place pour la vie privée au sein de l'entreprise ?

Un consensus a cependant été observé concernant l'usage de l'IA : il doit s'agir d'un outil pour l'entreprise qui ne doit en aucun cas analyser ce qui relève de la sphère privée. L'enjeu est alors de

délimiter cette frontière afin d'identifier clairement quand l'IA peut être utilisée ou non (il s'agit de définir quelles sont les données qui relèvent uniquement de la sphère professionnelle et peuvent servir son analyse).

L'analyse comportementale réalisée par le SIA a également été dénoncée. Celle-ci pourrait porter atteinte tant à la vie privée qu'à l'autonomie (à savoir, les citoyens ont questionné par la suite s'il était éthique d'utiliser une IA pour retracer des conversations et comportements, même s'ils relèvent de la sphère professionnelle). Les citoyens dénoncent ici une forme d'intrusion et un bris de confidentialité issus de cette surveillance constante :

> « C'est comme le Big Brother qui nous surveille »
> Un participant

Certains des participants ont souligné que ces enjeux n'étaient pas spécifiques à l'IA, tandis que d'autres considèrent que la haute traçabilité permise par l'IA les potentialise.

BIEN-ÊTRE

L'impact de ce genre de système sur le bien-être des employés a été considéré à la fois comme positif et négatif. En effet, si cette technologie peut être utilisée pour aider l'employé et améliorer la qualité de sa vie professionnelle (en participant à apaiser les relations, à déceler les cas de harcèlement ou d'intimidation ou encore en participant à la prévention du suicide), il semble qu'elle puisse tout autant lui nuire (le « déstabiliser », le rendre « inquiet », « normaliser » ses comportements). Les participants s'entendent sur un excès de sollicitation de l'employé dans le scénario, qui ne devrait pas avoir à justifier tous ses agissements. Les citoyens soulèvent ici un dilemme entre le bien-être et la liberté des employés. Jusqu'à quel point est-il possible de surveiller leurs gestes dans la perspective de protéger leur bien-être sans nuire indûment à leur liberté d'agir ? Qu'est-ce qu'une surveillance « bien utilisée » ? Les citoyens ont également souligné une corrélation entre le bien-être et les objectifs de performance : un employé heureux est également plus productif. Protéger le bien-être semble donc aller dans le sens des intérêts de l'entreprise comme de l'individu.

TRANSPARENCE

Les citoyens ont d'abord soulevé le non-respect du consentement de l'employé, qui ignore qu'il est « surveillé », et en appelle à plus de transparence de l'employeur qui devrait l'informer sur ce qui est enregistré ou non (notamment, dans le cadre de la formation des employés).

> « Le consentement de l'employé est important dans le recueil de données. Il faut être transparent envers ses employés. »
> Un participant

Cet enjeu de transparence a soulevé de nombreuses questions : Quelles sont les règles des rapports relationnels dans l'entreprise ? Quelle est la hiérarchie dans l'importance des données ? Qui peut les voir ? Un employé a-t-il le droit de voir les données de son patron ?

L'enjeu de transparence a également trait à l'IA ou au comment la rendre interprétable. Les citoyens ont soulevé ici la nécessité de communiquer autour du fonctionnement de l'algorithme afin notamment d'assurer une « adhésion » des individus à ces nouveaux systèmes. Ils ont également souligné l'importance de ne pas prendre une décision sur la base de la conclusion d'une technologie d'IA non explicable.

PRODUCTIVITÉ ET PERFORMANCE

L'enjeu de la performance et de son évaluation a ensuite été soulevé à plusieurs reprises. À quel point l'IA doit intervenir dans le but d'augmenter la productivité de l'employé (ex. en l'empêchant de se répéter dans les discussions de groupe) ? Est-ce essentiel ? Il serait ici nécessaire de définir plus précisément comment l'IA pourrait mesurer et améliorer la performance. D'une part, le risque de valoriser la productivité au détriment de l'employé,

de son développement personnel et de son comportement au travail a été soulevé. L'employé a-t-il vraiment une chance de se perfectionner dans ce contexte ? Faut-il personnaliser les attentes et les objectifs selon les personnes ?

« L'IA fait oublier la personne »

Un participant

D'autre part, la définition même de ce qui doit être considéré comme productif (et contre-productif) s'avère ici essentielle pour les participants. Quels indicateurs utiliser ? Quelle est leur pertinence ? Qui peut et doit les définir ? Doivent-ils être en lien avec les objectifs de l'entreprise ? Ces objectifs doivent-ils être révisés à la lumière des analyses de l'IA ? Bien qu'elle soit utilisée dans l'optique de vérifier que les employés « performent », l'IA, en interprétant les résultats des employés, pourrait tout aussi bien s'avérer être un frein à l'innovation, notamment parce que certaines tâches sont plus faciles à mesurer que d'autres.

L'ensemble des débats concernant la définition de la performance, de la productivité et du respect du bien-être des employés a mené à la conclusion que ces enjeux sont très liés aux cultures des entreprises, qui peuvent être très différentes et refléter des objectifs, des intérêts et des valeurs variés.

Certains citoyens autour de la table s'inquiètent que l'instauration d'outils basés sur l'IA pousse les sociétés à imposer une performance systématiquement associée à un « score », qui pourrait se traduire par une normalisation. D'autres mentionnent que ces pratiques existent déjà, mais qu'elles seront seulement amplifiées, voire « industrialisées » par les systèmes d'IA qui peuvent traiter un plus grand nombre d'informations beaucoup plus rapidement. D'autres encore ont mentionné qu'une uniformisation des pratiques peut tout de même être utile, notamment pour répondre aux besoins de l'entreprise.

Un débat sur la prétendue objectivité de l'IA a ensuite eu lieu, soulevant des questions telles que : Comment connaitre les critères de décision et d'automatisation ? De quelle manière l'IA va définir une absence d'activité de l'employé ? Comment équilibrer d'un côté la performance, l'objectivité et la standardisation par les technologies d'IA et de l'autre la subjectivité, la particularité et l'arbitrage humain ?

Après plus d'une heure de discussion, les participants ont sélectionné parmi ces cinq enjeux les trois qui semblaient prioritaires pour 2025. Ces enjeux font parfois écho aux principes de la version préliminaire de la Déclaration.

Tableau 5 : Les enjeux prioritaires

Enjeux éthiques 2025	**1**	**2**	**3**
Description	Comment encadrer l'évaluation de la performance en respectant à la fois les objectifs de l'entreprise (productivité) et l'individu (normalisation) ?	Comment faire en sorte qu'entreprise et employés comprennent l'IA (et garantir l'adhésion) ?	Comment l'IA peut préserver (contribuer, soutenir) le bien-être des employés.
Principes associés	Performance	Transparence (connaissance)	Bien-être

Deuxième moment délibératif : proposition d'encadrement de l'IA pour 2018-2020

Lors de cette deuxième partie de l'activité, les participants ont été sollicités pour formuler des recommandations et imaginer quelles solutions pourraient être mises en place pour répondre à ces trois enjeux. Une grande diversité de mécanismes plus ou moins contraignants a été formulée pour chacun des enjeux identifiés, pour finalement converger en six principaux mécanismes à mettre en place pour tenir compte de l'ensemble des enjeux.

Pour répondre à l'enjeu de **performance,** les participants ont d'abord recommandé d'organiser des **formations continues** dans les entreprises afin d'accompagner les gens dans chaque étape de la « transformation numérique ». L'**éducation** au numérique permettrait d'encourager une ambiance d'apprentissage, mais également de diminuer la peur qui peut accompagner le développement de l'IA dans le monde du travail.

> « Des formations continues à chaque étape de la transformation numérique d'une entreprise pour tous les salariés (encourager une ambiance d'apprentissage constant). »
> Un participant

Également, la mise en place d'une **autorité administrative indépendante** (AAI) et d'**un correspondant** au sein de l'entreprise a été envisagée comme pistes de solution[2], afin notamment de garantir le respect du dispositif RGPD (qui devrait être étendu à la traçabilité des données et l'explicabilité des décisions algorithmiques). Ce correspondant serait responsable de l'application des règlements, de soutenir le plaignant lors de problème et pourrait faire appel à l'AAI si besoin. Les participants ont également proposé la création **d'indicateurs** en lien direct avec non seulement les objectifs de l'entreprise (ex. résultats financiers), mais aussi avec le respect de certaines « valeurs humaines » (ex. bien-être du salarié). Pour cela, la mise en place d'une loi est absolument nécessaire selon les citoyens, si l'on ne veut pas que ce score soit uniquement corrélé aux intérêts financiers de l'entreprise.

Il a également été recommandé que le gouvernement crée un **programme de recherche publique** sur l'*interprétabilité* et la transparence des algorithmes (pour rattraper le retard sur la recherche privée à ce niveau), afin de comprendre comment ces décisions sont prises et de limiter le monopole des grosses entreprises d'IA. Dans la même optique, la création de « **groupes pilotes** » (ou « groupes tests ») afin de mesurer les impacts (incluant les impacts psychologiques et sociaux) de l'utilisation de l'IA au sein des entreprises et en vérifier la pertinence et l'utilité devraient être mis en place.

Pour répondre à l'enjeu de **transparence**, certains participants ont défendu la mise en place de mesures moins contraignantes, comme une **communication obligatoire** sur les différentes règles suivies, les données des salariées qui sont utilisées, les objectifs de leur collecte (scoring individuel ou collectif ?) et ce qu'on est capable de déduire de ces données, ou encore les résultats des analyses faites sur les groupes pilotes mentionnés précédemment. Cette communication doit être réalisée auprès de tous les départements (RH, informatique, marketing, juridique), et inclure des notions sur l'IA, une formation sur ce qu'est un algorithme et comment il apprend à partir des données.

Pour protéger les enjeux de **bien-être**, les participants ont recommandé **une évaluation annuelle** de la perception des employés en lien avec l'usage des dispositifs d'IA. Cette évaluation pourrait éventuellement être consolidée par un comité (par exemple le CHSCT), responsable d'agir en cas de problèmes. Celle-ci devrait favoriser les bonnes relations entre salariés et employeurs, et assurer que le mode de travail de chacun est respecté même avec l'avènement de la technologie. La mise en place d'**une certification** (ou label) qui serait garante de bonnes pratiques au point de vue éthique, environnemental et sociétal devrait être développée et imposée par l'État. Cette certification permettrait de garantir le respect d'un minimum de critères au sein des entreprises, afin d'optimiser la productivité

[2] Règlement général sur la protection des données, règlementation européenne entrée en vigueur le 25 mai 2018.

et la performance. Les **indicateurs** de bien-être, au même titre que ceux de la performance, devraient être pris en compte.

En réponse aux enjeux de bien-être et de transparence, les citoyens ont proposé la mise en place d'une loi qui doit définir et imposer l'*interprétabilité* des IA (notamment, la justification de la décision et un accès garanti à des règles explicites). Cette loi devrait dicter les critères minimums pour protéger le bien-être des individus (incluant un droit à la déconnexion) afin de garantir la protection des droits fondamentaux qui pourraient être menacés par le développement de l'IA.

Pour tous ces enjeux, et dans l'optique de « règlementer sans pénaliser », devrait être créée une **charte**[3] officielle qui reprend les droits, les devoirs et les valeurs à défendre pour protéger l'individu au sein de l'entreprise.

Au final, **6 recommandations** reprenant les grandes lignes des propositions précédentes ont été formulées de manière consensuelle :

Tableau 6 : Les propositions retenues

Propositions d'encadrement	**1**	**2**	**3**	**4**	**5**	**6**
Description	Loi qui doit définir et imposer l'*interprétabilité* des IA et qui définit les critères minimums pour protéger le bien-être des individus.	Certification (ou label).	Formation pour les différents acteurs de l'entreprise.	La mise à jour du Code du travail afin de l'adapter à la réalité numérique.	La création d'une charte des droits et devoirs.	Le financement de recherches publiques et d'études pilotes sur l'IA et ses impacts sur le monde du travail.
Enjeux associés	*Les 6 recommandations ont été formulées pour que chacune réponde aux 3 enjeux prioritaires*					

[3] Les participants ont souligné qu'une charte n'a pas la même force contraignante en France qu'au Québec.

Troisième moment délibératif : écriture de la « Une » d'un journal en 2020.

Cette étape est la mise en récit d'une des pistes de solution proposées en 2020. Les participants ont ici mis en scène un des risques de l'utilisation de l'IA au sein des entreprises et une des mesures envisagées pour y répondre.

Le premier employé renvoyé à cause de l'IA

Un salarié se fait renvoyer après trois semaines de travail sur les recommandations d'une IA. Il a contesté son renvoi auprès des Prud'hommes. Le jugement a été rendu après le débat à l'Assemblée. Une loi encadrant cette pratique va être votée.

Les participants ont tenu à souligner qu'ils ont surtout discuté des impacts potentiellement négatifs de l'IA en milieu de travail. Cependant, ils ont reconnu que l'IA pourrait également apporter de nombreux avantages tant pour les employés que les entreprises, avantages qui pourraient faire l'objet d'une autre discussion. La mise en place d'un organisme interne appuyé par un mécanisme législatif semble être la solution indispensable au développement responsable de l'IA au sein des entreprises.

3. DISCUSSION AUTOUR DU THÈME DE LA CULTURE avec les membres de la Coalition de la diversité des expressions culturelles (CDEC)

Afin d'aborder les enjeux relatifs au développement de l'IA dans le domaine de l'art et de la culture, un atelier de discussion a été organisé avec la Coalition de la diversité des expressions culturelles (CDEC) le 25 septembre 2018, réunissant près de 11 experts et parties prenantes du domaine. La discussion avec les participants s'est déroulée successivement autour de trois thématiques :

1. Les droits d'auteur ;
2. L'enjeu de la diversité culturelle ;
3. La propagande et la manipulation.

Suite à ces discussions, la CDEC a produit un mémoire[4] particulièrement pertinent qui fait état des différents défis et opportunités liés au développement de l'IA en culture. Ce mémoire présente également les principes éthiques nécessaires au développement responsable de l'IA dans ce domaine et les principales recommandations qui ressortent de la rencontre du 25 septembre. La présente section résume ces discussions et les points principaux de ce mémoire.

3.1.

TROIS THÉMATIQUES PROPOSÉES PAR L'ÉQUIPE DE LA DÉCLARATION POUR ABORDER LES ENJEUX DU DÉVELOPPEMENT DE L'IA DANS LE DOMAINE DE LA CULTURE

LES DROITS D'AUTEUR DANS UN CONTEXTE DE COCRÉATION PAR LES IA

Le recours à des systèmes d'IA pour générer des œuvres à très bas prix (ex. par les *Generative Adversarial Network*) en musique, arts visuels, séries télévisuelles, ou pour la rédaction d'articles de journaux, va reposer de manière inédite la question des droits d'auteurs. Doivent-ils revenir aux auteurs qui ont créé les exemples sur lesquels apprennent les algorithmes, ou bien aux informaticiens programmeurs d'algorithme, ou au promoteur du projet ? Le pastiche ou le *remix* produit par un algorithme génératif est-il du plagiat ? Si l'IA remplace l'artiste, quel sera l'effet sur la diversité culturelle ? Et si une IA « interprète » une œuvre ? Quel est l'accès au patrimoine artistique par les algorithmes ?

Cette problématique concerne surtout les SIA générateurs de création (voir les applications de l'IA dans le domaine culturel, Mémoire de la CDEC, p. 2).

L'ENJEU DE LA DIVERSITÉ CULTURELLE FACE AUX NOUVEAUX DISPOSITIFS DE RECOMMANDATION PAR ALGORITHME

Face aux risques d'uniformisation des goûts et des conduites, de « bulle algorithmique », de capture de l'attention et de formatage des choix par les algorithmes de recommandation, comment maintenir la diversité de l'offre culturelle ? Quelles pratiques de réception libre, autonome et critique promouvoir auprès des usagers ? Quelles possibilités de déconnexion (objets connectés, robots

[4] Pour plus d'informations sur les enjeux du développement de l'IA sur la diversité culturelle, consulter le mémoire : https://cdec-cdce.org/des-principes-ethiques-pour-un-developpement-de-lintelligence-artificielle-misant-sur-la-diversite-des-expressions-culturelles/

domestiques) face aux stratégies de capture de l'attention ? Quelles transparence et explicabilité pour les usagers ? Une politique culturelle publique devrait-elle être proposée sur cet enjeu de la diversité ? Avec quels nouveaux mécanismes de financement de la diversité culturelle ?

Cette problématique concerne surtout les algorithmes de recommandation et la valorisation des données (voir les applications de l'IA dans le domaine culturel, Mémoire de la CDEC, p. 2).

LA CENSURE ALGORITHMIQUE ET L'ART ACTUEL

Les algorithmes de reconnaissance sont utilisés par les médias sociaux pour exercer un pouvoir de censure parfois jugé excessif. Des artistes contemporains répondent à ces dispositifs par des interventions pour à la fois rendre visibles ces règles et les détourner. Quelle liberté critique pour les artistes contemporains de demain ? Quelle formation des artistes pour leur donner une connaissance critique ?

Cette problématique concerne surtout les algorithmes de recommandation et la valorisation des données (voir les applications de l'IA dans le domaine culturel, Mémoire de la CDEC, p. 2).

3.2

LES ENJEUX DE LA PROMOTION DE LA DIVERSITÉ CULTURELLE À L'HEURE DE L'IA

DÉCOUVRABILITÉ ET HOMOGÉNÉISATION DES CONTENUS CULTURELS

« Lorsque les recommandations sont basées, entre autres, sur la popularité des contenus, elles contribuent fortement à la concentration des écoutes (0.7 % des titres représentent 87 % des écoutes sur les services de musique en ligne au Canada), favorisant une minorité d'artistes »
(Mémoire de la CDEC, p. 4).

La **découvrabilité** a été identifiée comme un enjeu important de l'avènement de l'IA dans le domaine de la production culturelle. D'un côté, s'ils sont correctement paramétrés, les SIA pourraient devenir des outils de diversité culturelle, en élargissant par exemple les auditoires mondiaux. Ils pourraient proposer une diversité de contenu relativement élevée, en permettant par exemple à des artistes de se faire connaitre sans nécessairement passer par des intermédiaires, voire sans coût de production. D'un autre côté, les participants s'inquiètent d'un risque d'uniformisation des contenus culturels et de la réelle neutralité du web. Par exemple, les algorithmes de recommandations proposent rarement du contenu québécois sur les différentes plateformes culturelles comme Netflix, Amazon, ou Spotify. Ces plateformes comportent un très grand nombre de contenus, si bien que les utilisateurs s'en remettent souvent à l'algorithme de recommandation pour faire leur choix. Les participants soulèvent le risque que les algorithmes finissent par enfermer les individus « dans un goût », empêchant de découvrir d'autres contenus que ceux recommandés sur la base de leur choix précédent. Ceci risque d'entrainer une homogénéisation des contenus culturels,

notamment parce que les créateurs, pour la plupart, adaptent leurs œuvres à ce mode de diffusion.

Les participants soulignent qu'il est nécessaire que le milieu audiovisuel fasse des efforts de découvrabilité dans les plateformes numériques. Le problème ne vient pas seulement des SIA, c'est aussi un enjeu lié aux objectifs de l'industrie. Si les gouvernements ont une responsabilité concernant la diversité culturelle, les multinationales viennent toutefois défier leur pouvoir. Le Canada a par exemple été le premier signataire de la Convention de l'UNESCO pour la protection et la promotion de la diversité des expressions culturelles. Bien que certains contenus soient financés publiquement, le modèle dominant reste un modèle commercial, ce qui peut être problématique, comme lorsqu'on constate que les livres québécois ne sont pas mis de l'avant par l'algorithme d'Amazon (sur Amazon.ca), et ce malgré la volonté provinciale.

Les participants ont ainsi insisté sur le développement d'une démocratisation du pouvoir de créer et d'utiliser les SIA en ce sens, qui deviennent alors des agents importants de l'écosystème culturel en ce qui a trait à l'émergence des créations.

LES CONSÉQUENCES SUR L'EMPLOI

« En mai 2018, des chercheurs révélaient les résultats d'une enquête menée auprès de plus de 350 chercheurs en IA. En moyenne, ils prédisent que l'IA pourra dépasser les humains pour produire des essais de niveau scolaire en 2026, des chansons populaires en 2028 et des best-sellers en 2049. »
(Rapport de la CDEC, p. 3)

La possibilité pour une IA de générer une œuvre à part entière sans impliquer d'artiste est un sujet d'inquiétude. Par qui ces œuvres seront-elles développées ? Uniquement par des entreprises qui possèdent déjà un certain capital ? Est-ce compatible avec le développement d'une société composée de plus d'artistes et plus de diversité ? Les participants ont ici dénoncé l'existence d'un risque de boucle fermée (notamment en lien avec les travaux sur la directive européenne à ce sujet), où seulement une infime minorité d'artistes serait favorisée par les algorithmes. Ils craignent que le nombre d'artistes décline de façon importante.

Comment les créateurs humains pourraient-ils se démarquer face aux créations logicielles ? Comment cela va-t-il affecter les capacités individuelles et collectives de création et d'innovation ? Dans une expérimentation conduite chez l'ACTRA (Alliance of Canadian Cinema, Television and Radio Artists), il n'a pas été possible de faire la différence entre les personnages créés par une IA et ceux de vrais humains. Cela semble aller au-delà des enjeux du *DeepFake*, puisqu'il s'agit de la création de nouveaux visages. Or c'est un des rôles d'un grand nombre de leurs membres, qui travaillent dans le jeu vidéo. L'enjeu soulevé ici est celui de la perte d'emploi que ces technologies pourraient engendrer.

La question de la rémunération a également été soulevée, ouvrant la voie à une discussion sur les droits d'auteur. Qui rémunérer et comment, si des algorithmes sont à l'origine de la création ? La rémunération des acteurs se calcule selon des journées de travail et les droits de suite. Au niveau des droits de suite, des compromis sont envisageables, comme des normes pour réguler la réutilisation des œuvres. Les participants ont également souhaité tempérer l'impact potentiel de l'IA sur la reconnaissance de l'artiste, qui pourrait être plus limité que prévu, comme pour le livre numérique qui n'a pas tué le livre papier.

REPENSER LES DROITS D'AUTEUR

« Évidemment, la dématérialisation des contenus culturels, les changements technologiques et l'arrivée de nouveaux joueurs qui ont ***transformé les modèles d'affaires*** *ont un impact majeur sur la* ***rémunération des artistes,***

et le paiement de redevances de droits d'auteur. »
(Mémoire de la CDEC, p. 5)

Différentes questions ont été soulevées à ce sujet : Quelles seront les redevances ? Quelles seront les régulations concernant les droits d'auteur ? Quelle portion d'un manuscrit est nécessaire à un algorithme pour en produire un autre ? De quelle manière l'IA va pouvoir se baser sur d'autres livres pour en commercialiser de nouveaux ? L'avènement de SIA créateurs soulèvent ainsi d'importants enjeux de propriété des contenus culturels. Concernant la musique, la grande puissance de calcul et le plus grand nombre de données disponibles, combinées aux avancées majeures concernant l'efficacité des algorithmes, a permis la création d'outils de « génération » artistique. En musique, généralement, l'apprentissage se fait cependant déjà en reprenant ce qui a été fait. Si l'IA fait de même, c'est cependant à une vitesse incomparable.

Les participants ont ainsi reconnu que l'IA bafoue déjà les droits d'auteurs et ont remis en question la possibilité d'accorder ce genre de droits à une machine. Considérant que les satires et les parodies font déjà partie d'exceptions au droit d'auteur, serait-il possible de faire également exception pour l'IA ? La loi concernant les droits d'auteur est actuellement en révision. Il semble ici essentiel pour les participants qu'elle intègre les évolutions liées aux SIA dans sa nouvelle forme.

3.3

LES PRINCIPES ÉTHIQUES DE LA CDEC

Dans son mémoire, la CDEC recommande l'adoption de quatre principes éthiques pour prévenir les dérives du développement de l'IA dans le milieu culturel. Le respect de ces principes devrait permettre, selon eux, que l'IA intègre mieux les enjeux culturels en général et ceux de la diversité en particulier. Ces quatre principes sont cohérents avec ceux développés dans la version finale de la Déclaration de Montréal, mais intègrent les spécificités du domaine culturel et du domaine des arts.

La CDEC propose **un principe de diversité des expressions culturelles** qui fait directement écho au **principe d'inclusion de la diversité.** Cependant, la CDEC précise ici que ce principe devrait faire en sorte que les SIA « valorisent les contenus culturels et linguistiques locaux au sein des populations dont ils sont issus, favorisant ainsi la cohésion sociale tout comme le tissu économique local ; incitent les usagers à faire des découvertes à l'extérieur de leur univers ; facilitent le passage entre les familles technologiques (ex. Apple), plutôt que de les y enfermer ; favorisent les interactions et le partage des contenus. » (Mémoire de la CDEC, p. 6)

La CDEC propose également **un principe de valorisation de la culture, des artistes et des producteurs** de contenus culturels, soit que la contribution de l'IA permette d'éviter l'actuelle dévalorisation des contenus culturels et qu'ils ne « favorisent l'appropriation démesurée des revenus qui devraient être destinés aux écosystèmes culturels. » (Mémoire de la CDEC, p. 6). Si ce principe est en lien avec plusieurs des principes de la Déclaration, il appelle surtout à respecter le 6e principe d'équité et les sous-principes associés : *Le développement et l'utilisation des SIA doivent contribuer à la réalisation d'une société juste et équitable.*

La CDEC propose ensuite un **principe de transparence** (au niveau du code des algorithmes, mais aussi des données prises en compte) **et de dialogue** (notamment, avec les utilisateurs). Ce principe appelle à respecter le 5e principe de la Déclaration, celui de participation démocratique : Les SIA doivent satisfaire les critères d'intelligibilité, de justifiabilité et d'accessibilité, et doivent pouvoir être soumis à un examen, un débat et un contrôle démocratiques. De même pour le 2e principe de la Déclaration, celui d'autonomie : *Les SIA doivent être développés et utilisés dans le respect de l'autonomie des personnes et dans le but d'accroitre le contrôle des individus sur leur vie et leur environnement.*

Enfin, la CDEC propose un principe de **primauté de l'intérêt public,** défini comme suit : *Toutes les innovations technologiques ne sont pas désirables. Le développement de l'IA devrait toujours privilégier l'amélioration de la qualité de vie des populations, de la cohésion sociale et des pratiques démocratiques. Les gouvernements doivent défendre l'intérêt public face à des développements qui pourraient avoir des impacts plutôt négatifs sur la société.* (Mémoire de la CDEC, p .7). Le respect de ce principe s'accorde avec le respect du 1er principe de la Déclaration, le principe de **bien-être** : *Le développement et l'utilisation des systèmes d'intelligence artificielle (SIA) doivent permettre d'accroitre le bien-être de tous les êtres sensibles*. Et également avec le 8e principe de la Déclaration, celui de **prudence** : *Toutes les personnes impliquées dans le développement des SIA doivent faire preuve de prudence en anticipant autant que possible les conséquences néfastes de l'utilisation des SIA et en prenant des mesures appropriées pour les éviter.*

Principes éthiques de la CDEC	Principes de la Déclaration de Montréal
Principe de diversité des expressions culturelles	Principe 7 d'inclusion de la diversité
Principe de valorisation de la culture, des artistes et des producteurs de contenus culturels	Principe 6 d'équité
Principe de transparence et dialogue	Principe 2 du respect de l'autonomie Principe 5 de participation démocratique
Principe de primauté de l'intérêt public	Principe 1 de bien-être durable Principe 8 de prudence

3.4

QUELQUES-UNES DES RECOMMANDATIONS FORMULÉES

Différentes recommandations sont ressorties des discussions du 25 septembre. Elles ont été formulées dans le but de promouvoir le contenu culturel québécois et conscientiser les citoyens aux impacts du développement de l'IA sur la culture. D'abord, pour favoriser la diversité de l'expression culturelle dans l'univers numérique, les participants recommandent la définition d'exigences minimales de représentation de contenus culturels canadiens dans les recommandations issues d'algorithmes, comme c'est déjà le cas pour la télévision ou la radio québécoise. Ces exigences doivent être formulées dans des lois et règlementations, car elles ne pourront, selon les participants, émaner du libre marché.

Lors des discussions, les participants ont reconnu que la maîtrise de la littératie relative au développement de l'IA est essentielle. Il est nécessaire d'outiller les individus de manière à ce que chacun puisse comprendre où les recommandations des algorithmes les amènent. Les participants recommandent la mise en place d'une politique d'éducation des usagers, qui aurait pour but de contrer la fausse impression de choix, en les incitant à varier leur exploration et rester conscients de l'influence des algorithmes. Cette politique irait dans le sens d'une éducation au choix critique. Ainsi, une forme d'autodéfense intellectuelle devrait être développée pour tous, et ce dès l'enfance. Les participants recommandent également de sensibiliser les développeurs informatiques aux impacts des SIA sur la culture.

Lors de ces discussions est également ressortie une recommandation concernant la transparence et l'explicabilité des recommandations algorithmiques. Les usagers devraient être systématiquement informés lorsqu'une recommandation émane d'un SIA et devraient avoir accès facilement à de l'information explicative tant sur le fonctionnement des algorithmes que sur l'existence d'autres contenus culturels.

Les participants ont également recommandé que les entreprises qui développent des SIA ayant un impact sur la culture consacrent une part de leur chiffre d'affaires à la promotion de la diversité culturelle, en finançant par exemple, certaines bibliothèques, évènements culturels, ou médias. Les participants prônent également la mise en place d'une surveillance sur le profilage de goût et sur la protection des renseignements personnels. Ils suggèrent aussi le développement d'un « laboratoire de l'IA en culture » afin d'observer les algorithmes, d'apprendre à interagir avec eux et éventuellement d'influencer leur développement.

Deux principales recommandations issues de ces discussions sont particulièrement développées dans le mémoire de la CDEC :

1. Éducation et formation
2. Révision des lois touchant le milieu culturel

Ces recommandations s'inscrivent dans la lignée de celles formulées pour les autres secteurs lors de la coconstruction de l'hiver, à savoir 26 % des recommandations comptabilisées sont des dispositions légales et juridiques et 19 % des formations (*cf. Rapport des résultats des ateliers de coconstruction de l'hiver*).

4. FAIRE LE PONT ENTRE LES DÉLIBÉRATIONS CITOYENNES ET LA RELÈVE EN RECHERCHE : SIMULATION DE LA RÉDACTION DE BRÈVES POLITIQUES

4.1.

DESCRIPTION DE L'ACTIVITÉ

Afin de faire le pont entre la relève en recherche et les citoyens, la Déclaration a participé à l'organisation d'une activité de simulation en partenariat avec le Comité intersectoriel étudiant (CIÉ) des Fonds de recherche du Québec (FRQ) et l'École de politique appliquée (EPA) de l'Université de Sherbrooke dans le cadre des Journées de la relève en recherche (J2R) organisées par l'Acfas. L'activité de simulation « Politique et intelligence artificielle » avait pour but de réunir des étudiants de la relève en recherche dans l'optique de produire trois brèves politiques sur l'IA. L'objectif de cette activité était de permettre à la relève de prendre part aux discussions relatives à l'IA et aux enjeux éthiques et sociaux de son développement. Ce thème s'est avéré fédérateur pour le CIÉ :

> « L'IA a été choisie car elle présente des dimensions intersectorielles et englobe des enjeux intéressants d'un point de vue de l'alliage entre science, société et développement de politiques publiques. Dans le cadre de cette simulation, l'IA permet à la relève de prendre part aux réflexions et au positionnement du Québec comme un leader en cette matière. »
> (Guide à l'intention des participants, p. 5).

Dans cette optique, la Déclaration de Montréal a fourni trois problématiques issues des discussions de la coconstruction citoyenne réalisée lors de l'hiver 2018. Il nous a semblé pertinent, tant pour l'élaboration des travaux de la Déclaration, que pour le travail des jeunes chercheurs sélectionnés pour l'activité, de débattre et émettre des recommandations autour de ces thèmes portés par les délibérations. Ces trois problématiques touchent en effet à des enjeux particulièrement sensibles du développement de l'IA sur lesquels il est urgent de délibérer :

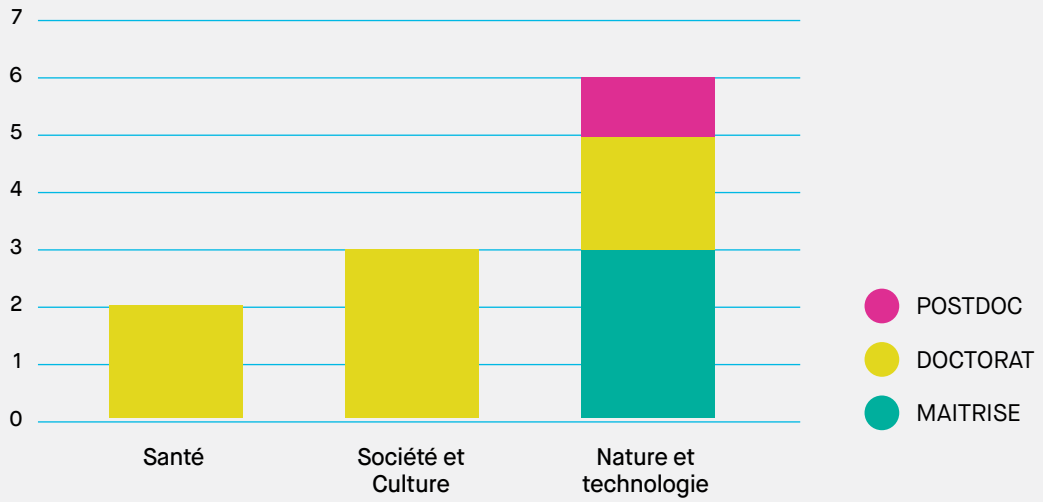
1. La sécurité et l'intégrité des systèmes d'IA : Comment maximiser les impacts positifs tout en minimisant les effets néfastes du développement de l'IA ?
2. L'IA, les médias et la manipulation de l'information : Comment lutter contre la diffusion et l'amplification de fausses nouvelles et de campagnes de désinformation ? Comment favoriser la démocratisation de l'accès à l'information tout en encourageant la formation d'une pensée critique et la prise de décision éclairée ?
3. La gouvernance publique, privée et participative ; les communs numériques : Laquelle de ces gouvernances serait la plus appropriée ? Quelles balises mettre en place ?

La simulation s'est déroulée en vue de trois objectifs : « 1. Familiariser les participants à l'écriture et à la présentation d'une brève politique ; 2. Faciliter l'acquisition de compétences liées au développement de politiques scientifiques ; 3. Analyser une problématique sociale avec un regard scientifique. » (Guide à l'intention des participants, CIÉ, p. 6). Les brèves politiques ont ainsi été élaborées dans le cadre d'un exercice se voulant avant tout pédagogique : la diffusion

auprès de décideurs ou parties prenantes n'était pas l'objectif de l'activité. Toutefois, les recommandations présentées dans ces brèves nous donnent accès à la perspective des étudiants de la relève ayant participé, et elles ont été particulièrement pertinentes. Les brèves sont ainsi annexées au présent rapport bien qu'elles ne soient pas porteuses de recommandations concrètes en vue de politiques publiques. Nous avons décidé de présenter les brèves telles que rédigées par les étudiants, sans révision, afin de ne pas travestir leurs productions et de représenter au mieux leur contribution (cf. Annexe 2).

L'activité s'est déroulée les 18 et 19 octobre 2018 à l'Université de Sherbrooke et a réuni 11 étudiants aux expertises et aux niveaux variés.

Graphique 5 : Profils des étudiants participants à l'activité en fonction de leur domaine d'étude (selon les secteurs des trois Fonds)



Les étudiants ont été répartis en trois groupes, chacun traitant d'une des problématiques et accompagné d'un facilitateur (indépendant de l'équipe de la Déclaration afin de ne pas orienter les recommandations). Après quatre présentations introductives sur l'IA et la rédaction de brèves politiques ; les étudiants ont disposé de six heures seulement pour rédiger les brèves et préparer une présentation orale de leur travail. Les trois groupes ont présenté leurs brèves devant un jury[5] le 19 octobre au matin. L'équipe 2 (traitant de la problématique L'*IA, les médias et la manipulation de l'information*) a remporté le concours.

[5] Le jury était composé de trois membres :

Claude Asselin, professeur, Département d'anatomie et de biologie cellulaire, Faculté de médecine et des sciences de la santé, Université de Sherbrooke. [Représentant de l'Acfas].

Benoit Sévigny, directeur du Service des communications et de la mobilisation des connaissances. [Représentant FRQ].

Nathalie Voarino, candidate au doctorat en bioéthique, coordonatrice scientifique de la Déclaration de Montréal. [Représentante de la Déclaration]

4.2

LES PROBLÉMATIQUES ISSUES DES PRÉOCCUPATIONS CITOYENNES

Problématique 1 : Sécurité publique et intégrité des systèmes

Lors des consultations, les citoyens ont reconnu que l'essor de l'IA pourrait contribuer à rendre nos environnements physique et numérique plus sécuritaires. Par exemple, dans une ville intelligente, l'autonomisation des transports pourrait réduire le taux d'accidents routiers ; en santé publique des modèles épidémiologiques pourraient permettre aux autorités de mieux prévoir la propagation des maladies ; en matière de cybersécurité, des spécialistes en sécurité informatique se tournent vers l'intelligence artificielle pour reconnaître les agressions.

Cependant, les citoyens ont reconnu que certaines conditions sont nécessaires pour s'assurer que les avancées de l'IA soient bénéfiques pour la sécurité publique. Garantir un « bon » usage de l'IA par le biais de l'intégrité et de la sécurité des systèmes sont des enjeux fondamentaux du développement responsable de ces technologies.

Les impacts négatifs de l'IA sur la sécurité publique peuvent prendre quatre formes :

1. **Une IA conçue dans le but de menacer la sécurité publique**[6]. Par exemple, l'utilisation de l'IA dans des buts de cybercriminalité (vol d'identité, *hacking* de centrales nucléaires, etc.), déstabilisation politique (propagande ciblée, création de fausses vidéos, etc.) ou automation d'équipement militaire (drones, soldats robots, etc.)[7]

2. **L'utilisation des informations collectées pour des fins autres que celles initialement prévues.** Ici, les citoyens craignent l'utilisation de dossiers médicaux complets par des compagnies d'assurance, celle de dossiers scolaires pour automatiser le marché de l'emploi, ou celle de systèmes d'optimisation de la circulation pour suivre et surveiller les usagers de la route.

3. **Un détournement volontaire de systèmes d'IA.** Quelqu'un de mal intentionné pourrait cibler directement le mode de fonctionnement de l'algorithme[8], par exemple, en trompant un système de reconnaissance faciale pour avoir accès à des données protégées. On pourrait aussi profiter des défis de sécurité liés à la prolifération des objets connectés[9], par exemple, pour prendre le contrôle d'une voiture autonome, ou pour paralyser un réseau avec une attaque massive de déni de service[10].

4. **Une mauvaise évaluation d'une IA.** Une intelligence artificielle dont la fiabilité ou la robustesse a été surestimée et qui se retrouve à l'origine d'un accident[11] [6]. Par exemple, les citoyens ont mentionné qu'un accident de camion autonome ou une erreur systématique d'un logiciel de diagnostic médical peut avoir de lourdes conséquences.

[6] Brundage M, Avin S, Clark J, Toner H, Eckersley P, Garfinkel B, et al. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. 2018;(February 2018). Disponible sur: http://arxiv.org/abs/1802.07228

[7] Bouvet M, Chiva E. Un regard (décalé ?) sur Intelligence Artificielle et Défense/Sécurité - CGE [Internet]. Conférence des Grandes Écoles. 2016 [cité le 3 sep 2018]. Disponible sur: http://www.cge.asso.fr/liste-actualites/un-regard-decale-sur-intelligence-artificielle-et-defensesecurite/

[8] Kurakin A, Goodfellow I, Bengio S, Dong Y, Liao F, Liang M, et al. Adversarial Attacks and Defences Competition [Internet]. 2018 [cité le 3 sep 2018]. Disponible sur: https://arxiv.org/pdf/1804.00097.pdf

[9] Zhang Z-K, Cho MCY, Wang C-W, Hsu C-W, Chen C-K, Shieh S. IoT Security: Ongoing Challenges and Research Opportunities. In: 2014 IEEE 7th International Conference on Service-Oriented Computing and Applications [Internet]. IEEE; 2014 [cité le 3 Sep 2018]. p. 230–4. Disponible sur: http://ieeexplore.ieee.org/document/6978614/

[10] Franceschi-Bicchierai Lorenzo. How 1.5 Million Connected Cameras Were Hijacked to Make an Unprecedented Botnet [Internet]. Vice Motherboard. 2016 [cité le 3 sep 2018]. Disponible sur: https://motherboard.vice.com/en_us/article/8q8dab/15-million-connected-cameras-ddos-botnet-brian-krebs

[11] Amodei D, Olah C, Brain G, Steinhardt J, Christiano P, Schulman J, et al. Concrete Problems in AI Safety [Internet]. [cité le 3 sep 2018]. Disponible sur: http://arxiv.org/abs/1606.06565.pdf

Les citoyens se sont donc demandé comment limiter les risques d'impact négatif de l'IA sur la sécurité (publique). Ils ont soulevé différents dilemmes potentiels en rapport avec la sécurité et l'intégrité des systèmes :

> Le respect de la transparence (un impératif souvent repris) pourrait-il nuire à la sécurité en facilitant le piratage ?

> Est-ce qu'assurer la plus grande sécurité possible implique forcément un compromis avec l'efficacité du système (qui doit être sécuritaire sans pour autant devenir inopérant) ?

> Et plus généralement, comment s'assurer de maximiser les impacts positifs du développement de l'IA tout en prévenant les effets néfastes ?

Autres références pertinentes :
Asilomar AI principles
Vidéo : Adversarial ML

Problématique 2 : L'IA, les médias et la manipulation de l'information

Les citoyens se sont inquiétés d'un risque de manipulation qui plane sur les utilisateurs à mesure que leurs actes sont de plus en plus influencés par des mécanismes d'IA qui influencent leurs décisions, souvent à leur insu ou via des incitatifs. Cela pose un problème de confiance en ces dispositifs puisqu'il y a une forme d'atteinte à l'autonomie et un risque d'orientation des actions (par exemple, en fonction d'intérêts privés). Les citoyens se demandent par exemple si les nouvelles technologies relevant de l'IA pourraient créer une nouvelle classe de lobby, qui parfois risquerait d'avoir trop de pouvoir. Pour entretenir une certaine liberté dans les choix orientés par l'IA et éviter d'accorder une confiance aveugle à ces dispositifs, il est donc important de cultiver une pensée critique chez tout citoyen et professionnel interagissant avec l'IA.

Si la propagande n'est pas un phénomène nouveau, la facilité et la vitesse avec lesquelles celle-ci peut être créée et diffusée, via des fausses nouvelles ou des campagnes de désinformation, sont sans précédent. Cela est notamment permis grâce aux plateformes de création et diffusion de contenu en ligne (via les réseaux sociaux, les blogues et sites internet, les forums de discussion) qui sont structurées sur des modèles de rétention de l'attention, de publicité et de recommandations[12-13-14-15]. Ce phénomène est également amplifié par la capacité de cibler très précisément des individus selon la collecte et l'analyse de leurs données personnelles comme l'a mis en lumière le scandale Cambridge Analytica[16], par exemple. La diversité des contenus à laquelle chaque individu est exposé est alors réduite à l'ensemble de ce qui se rapproche le plus de ce qu'il a déjà aimé, partagé, commenté. En se trouvant principalement confronté aux idées avec lesquelles il est en accord, l'individu est pris dans un effet de « bulle de filtre »[17] qui remet en question la capacité actuelle de tout citoyen à se forger une pensée critique.

Les grandes entreprises de médias sociaux annoncent des ensembles de mesures pour limiter le propagandisme potentiel de leurs outils (cf. les rapports de transparence de Facebook, Google et Twitter[18]). Mais est-ce suffisant ? Comment s'assurer que ces outils, qui ont démocratisé l'accès à l'information et la mise en relation, ne constituent

[12] Ingram M. Fake news is part of a bigger problem: automated propaganda [Internet]. Columbia Journalism Review. 2018 [cité 3 sept 2018]. Disponible à: https://www.cjr.org/analysis/algorithm-russia-facebook.php

[13] Lewis P. « Fiction is outperforming reality »: how YouTube's algorithm distorts truth. The Guardian [Internet]. 2 févr 2018 [cité 3 sept 2018]; Disponible à : https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth

[14] Marwick A, Lewis R. Media Manipulation and Disinformation Online [Internet]. Data & Society Research Institute; 2017 mai [cité 3 sept 2018]. Disponible à: https://datasociety.net/output/media-manipulation-and-disinfo-online/

[15] Tusikov N. Regulate social media platforms before it's too late [Internet]. The Conversation. 2017 [cité 3 sept 2018]. Disponible à: http://theconversation.com/regulate-social-media-platforms-before-its-too-late-86984

[16] Cadwalladr C, Graham-Harrison E. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian [Internet]. 17 mars 2018 [cité 3 sept 2018]; Disponible à: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[17] Pariser E. The filter bubble: what the Internet is hiding from you. London: Penguin Books; 2012.

[18] Rapport préliminaire de Facebook : https://transparency.facebook.com/community-standards-enforcement/
Rapport de transparence Google : https://transparencyreport.google.com/about
Rapport de transparence Twitter : https://transparency.twitter.com/fr.html

pas des outils de démocratisation de la propagande ? Comment lutter contre la diffusion et l'amplification de fausses nouvelles et de campagnes de désinformation afin de préserver la démocratie ? Comment favoriser la démocratisation de l'accès à l'information tout en encourageant la formation d'une pensée critique et la prise de décision éclairée ?

Autres références pertinentes :

Caplan R, Hanson L, Donovan J. Dead Reckoning, Navigating Content Moderation After Fake News. Data & Society Research Institute; 2018 févr [cité 3 sept 2018]. Disponible à: https://datasociety.net/output/dead-reckoning/

Foisy P-V. Facebook veut s'attaquer aux fausses nouvelles au Canada [Internet]. Radio-Canada.ca. [cité 3 sept 2018]. Disponible à: https://ici.radio-canada.ca/nouvelle/1109432/fake-news-facebook-fausses-nouvelles-canada-verification-faits

Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science. 9 mars 2018;359(6380):1094-6.

Jeangène Vilmer J-B, Escorcia A, Guillaume, M, Herrera J. Les manipulations de l'information, un défi pour nos démocraties [Internet]. Paris, France: CAPS et IRSEM; 2018 août. Disponible à: https://www.defense.gouv.fr/irsem/page-d-accueil/nos-evenements/lancement-du-rapport-conjoint-caps-irsem.-les-manipulations-de-l-information

Sites internet à consulter :

The Computational Propaganda Project : http://comprop.oii.ox.ac.uk

Observatory on Social Media : https://truthy.indiana.edu

Conversation AI : https://conversationai.github.io

"A Citizen's Guide to Fake News" sur le site du Center for Information Technology & Society, UC Santa Barbara : http://cits.ucsb.edu/fake-news

Problématique 3 : Gouvernance publique, privée ou participative : les communs numériques

À de nombreuses reprises, les citoyens ont soulevé des enjeux relatifs au partage de la gestion du développement de l'IA entre institutions publiques et privées, et les risques qui accompagnent ce partage, par exemple les conflits d'intérêts, la protection de l'indépendance des acteurs institutionnels ou des institutions publiques, la valeur marchande des données et la protection de la vie privée. Également, le risque de l'apparition d'un monopole privé dans la gouvernance du développement de l'IA a été mentionné à plusieurs reprises. En effet, en ce qui concerce notamment les compagnies propriétaires de données massives (qui sont à la base du fonctionnement de l'IA), certains s'inquiètent de l'apparition de monopoles, renforcés par les fusions de nouveaux fournisseurs de services plus petits[19].

En ce qui a trait à une gouvernance étatique, un encadrement légal et juridique de l'IA s'accompagne de différents risques et défis[20], par exemple, en se centrant trop sur les capacités des dispositifs aux dépens de la protection des valeurs humaines[21], si bien qu'on peut se demander s'il est possible de réglementer l'IA et remettre en question le pouvoir réel de l'État[22].

Si la discussion autour de la gouvernance oppose souvent les institutions publiques aux compagnies privées, une alternative a été proposée : celle d'une gouvernance participative qui donne directement la main aux citoyens en proposant, par exemple, la mise en place d'une grande consultation publique ou d'un espace permanent de concertation.

Dans cette perspective de gouvernance participative, la possibilité d'accorder de l'importance à la contribution des usagers dans la conception des outils d'IA et de leur gestion a ainsi

[19] IBig data: Bringing competition policy to the digital era - OECD [Internet]. [cited 2018 Sep 3]. Available from: http://www.oecd.org/competition/big-data-bringing-competition-policy-to-the-digital-era.htm

[20] Scherer MU. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. Harvard Journal of Law & Technology, Vol. 29, No. 2, Spring 2016. http://dx.doi.org/10.2139/ssrn.2609777

[21] Ambrose ML. Regulating the loop: ironies of automation law. 2014;38.

[22] Danaher J. Philosophical Disquisitions: Is effective regulation of AI possible? Eight potential regulatory problems [Internet]. Philosophical Disquisitions. 2015 [cited 2018 Sep 3]. Available from: http://philosophicaldisquisitions.blogspot.com/2015/07/is-effective-regulation-of-ai-possible.html

été proposée comme alternative. Cette participation pourrait prendre la forme d'une expérimentation collective (*design thinking*) par le biais de matériel en accès libre (*open source*). Ce matériel accessible à tous renvoie à la notion de (biens) communs numériques, soit l'ensemble des ressources et savoirs partagé et cocréé disponible en accès libre (par exemple, les logiciels libres). Plus qu'une simple forme de propriété, il s'agit ici d'un mode d'organisation coopératif garantissant l'horizontalité (échanges entre pairs) et la liberté d'expression[23]. Cette organisation dépend des formes de régulation décidées par les acteurs eux-mêmes.

« 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. »[24]

Ce mode de gouvernance n'est pas lui non plus sans défis, il est notamment fragile à différentes formes d'*enclosure* (réduction des usages communs) par l'État comme par les compagnies[25].

Ces enjeux soulèvent différents questionnements : quel serait le meilleur partage entre gestion publique, privée et participative pour la gouvernance de l'IA ? Ces différents modes entrent-ils forcément en tension et lequel serait le plus approprié ? Est-il nécessaire de baliser ces types de gestion et si oui, quelles balises mettre en place ?

Autres références pertinentes :

Chessen M. Encoded laws, policies, and virtues: the offspring of artificial intelligence and public-policy... [Internet]. Medium. 2017 [cited 2018 Sep 3]. Available from: https://medium.com/artificial-intelligence-policy-laws-and-ethics/encoded-laws-policies-and-virtues-the-offspring-of-artificial-intelligence-and-public-policy-3dfb357faf9

Shafto, P. Why Big Tech Companies Are Open-Sourcing Their AI Systems [Internet]. IFLScience. [cited 2018 Sep 3]. Available from: https://www.iflscience.com/technology/why-big-tech-companies-are-open-sourcing-their-ai-systems/

Sites internet à consulter :

Le site internet de la FACIL

La Déclaration des communs numériques de la FACIL

4.3.

LES RECOMMANDATIONS DE LA RELÈVE EN RECHERCHE

L'exercice s'est avéré particulièrement fructueux et les trois brèves sont porteuses de pistes de recommandations pertinentes en vue du développement responsable de l'IA.

La première brève politique traite des conséquences engendrées par une mauvaise évaluation des capacités de l'IA, un élément qui s'intègre dans la première problématique proposée sur la sécurité et l'intégrité des systèmes. Cette brève vise à promouvoir la sécurité et la protection des Canadiens, plus particulièrement en ce qui concerne les systèmes embarqués. Les recommandations énoncées se rapportent à la création d'un organisme pancanadien de certification des systèmes embarqués utilisant l'IA (ARBIA) et le développement d'un régime de responsabilité sans faute. Cette brève s'est démarquée par la pertinence de ses recommandations, d'une part en faisant référence à des mécanismes déjà existants (le MEI[26] et l'intégration des recommandations dans la Stratégie numérique du Québec) et, d'autre part, en proposant

[23] Crosnier HL. Communs numériques et communs de la connaissance. Introduction. tic&société. 2018 May 31;(Vol. 12, N° 1):1–12.

[24] Ruzé E. 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. 2013;(65):189–205.

[25] Crosnier HL. Une bonne nouvelle pour la théorie des biens communs. Vacarme. 2011;(56):92–4.

[26] Anciennement le MESI (Ministère de l'économie, de la science et de l'innovation), renommé MEI (Ministère de l'économie et de l'innovation) le jour même de l'écriture de la brève.

un mécanisme original de protection des citoyens (la responsabilité des dommages liés à un accident étant soit celle de l'État soit celle des entreprises qui développent lesdits systèmes embarqués).

La seconde brève, intitulée « Fausses nouvelles, vrais enjeux : s'éduquer pour y faire face » aborde le problème de la manipulation de l'information (en particulier, la création de fausses nouvelles) sur internet par le biais de l'IA. Les étudiants ont ici mis de l'avant la nécessité de favoriser le développement d'un esprit critique. Leurs recommandations vont dans le sens de l'importance de l'éducation face aux médias et à l'information au Québec. Deux principales recommandations ressortent de leur analyse : 1) Alerter la population québécoise sur la dissémination de fausses nouvelles (vigilance) notamment par le biais de SIA spécialisés et 2) Offrir des outils au corps enseignant québécois pour intégrer la conscientisation face aux fausses nouvelles à l'éducation dès l'école primaire (éducation). Afin d'assurer la mise en place de ces mécanismes, ils recommandent la création d'un Comité de vigilance numérique du Québec, annexé à l'*Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique et d'intégrer les processus éducatifs dans le Plan d'action numérique* du Québec.

La troisième brève traite de la problématique de la gouvernance de l'IA, explorant les enjeux liés aux brèches dans les politiques et les règlementations actuelles qui concernent le secteur de l'IA. Le défi est ici de trouver une forme de gouvernance de l'IA qui répond au mieux aux besoins des différents acteurs concernés (entreprises, citoyens, institutions publiques). L'objectif principal de cette brève politique est de *fournir une méthode de travail et de réflexion* afin de répondre adéquatement aux enjeux soulevés : politiques actuelles mal adaptées, inquiétudes des citoyens, flou relatif face à la responsabilité lors d'incidents et absence de méthodes pour gérer les problèmes liés à l'IA. Plusieurs pistes de solutions ont été proposées. Les étudiants recommandent la création d'un organisme (provincial) indépendant de régulation de l'utilisation des données communes qui serait responsable, entre autres, de la mise en place de mécanismes éducatifs ou, encore, d'ajuster les lois et réglementations en vigueur afin de les adapter aux nouvelles réalités technologiques. L'intégration d'un volet à la mission de l'organisme afin que celui-ci soit constamment en phase avec le marché est également proposée. L'organisme en question, nommé « Éduc'IA » lors de la présentation orale, s'organiserait sous la forme d'un *think tank*.

En résumé, les étudiants recommandent la mise en place d'un organisme indépendant de gestion de l'IA aux responsabilités variées (impliquant la création d'une certification ou d'un système de vigilance) et intégrant des processus éducatifs. Ces recommandations vont dans le sens de celles formulées lors des autres activités de coconstruction. Si la mise en place d'un organisme indépendant ou de mécanismes éducatifs s'aligne avec les recommandations de la Déclaration en vue de politiques publiques, celle de la création d'un système de responsabilité mériterait une analyse plus approfondie. Faisant écho aux recommandations citoyennes de l'hiver 2018 sur la mise en place de mécanismes assurantiels définissant les paramètres du partage des responsabilités en cas de faute (*cf. Rapport des résultats des ateliers de coconstruction de l'hiver*), cette recommandation invite à considérer une analyse à part entière portant sur la responsabilité civile et pénale face aux impacts de l'IA.

La pertinence de cette activité nous conduit enfin à appuyer le CIÉ dans sa recommandation de créer plus d'opportunités pour les étudiants aux cycles supérieurs de se former aux activités professionnelles extra-académiques, et plus particulièrement à la participation politique.

Ces trois activités nous ont ainsi permis d'explorer les enjeux liés au développement de l'IA à partir de nouvelles thématiques (ex. propagande) et également, d'expérimenter de nouvelles manières de procéder (ex. activité de simulation).

Quelle que soit l'activité, les recommandations des participants viennent appuyer la nécessité de mettre en place des formations adaptées pour tous et de mettre à jour le cadre légal et règlementaire lié à la progression de nouvelles connaissances sur le développement de l'IA et de ses impacts. Ces recommandations encouragent également la

promotion d'une gouvernance participative, mettant de l'avant l'importance d'impliquer les parties prenantes à différents moments clés de la gestion du développement de l'IA et de la décision politique.

En ouvrant la discussion sur des pistes de solution nouvelles ou des spécificités sectorielles, ces activités nous indiquent que la coconstruction mériterait d'être poursuivie au-delà des travaux de la Déclaration de Montréal, et renforcent la pertinence d'une consultation citoyenne sur le développement responsable de l'IA.

5. CONCLUSION

Ces trois activités nous ont ainsi permis d'explorer les enjeux liés au développement de l'IA à partir de nouvelles thématiques (ex. propagande) et également, d'expérimenter de nouvelles manières de procéder (ex. activité de simulation).

Quelle que soit l'activité, les recommandations des participants viennent appuyer la nécessité de mettre en place des formations adaptées pour tous et de mettre à jour le cadre légal et règlementaire lié à la progression de nouvelles connaissances sur le développement de l'IA et de ses impacts. Ces recommandations encouragent également la promotion d'une gouvernance participative, mettant de l'avant l'importance d'impliquer les parties prenantes à différents moments clés de la gestion du développement de l'IA et de la décision politique.

En ouvrant la discussion sur des pistes de solution nouvelles ou des spécificités sectorielles, ces activités nous indiquent que la coconstruction mériterait d'être poursuivie au-delà des travaux de la Déclaration de Montréal, et renforcent la pertinence d'une consultation citoyenne sur le développement responsable de l'IA.

ANNEXE 1
Les scénarios

DÉMOCRATIE

Fausse nouvelle dans la campagne électorale

23 mars 2022. Ce matin, Dominique B. se rend à la réunion de crise de l'Agence sur l'intégrité de l'information (AII), mise en place dans le cadre de la Loi contre la manipulation de l'information. Le président de la République sortant, candidat à sa réélection, vient de perdre 7 points dans les sondages d'intention de vote en trois semaines et la tendance à la baisse semble se confirmer. Alors qu'il était assuré de l'emporter deux mois auparavant, il est désormais dépassé par la candidate populiste de droite qui a pris la tête de la course électorale. Le tournant se situe le 2 mars, avec la diffusion sur internet d'une vidéo montrant le président de la République discuter avec le président du Mouvement des entreprises de France, en marge de son école d'été. Le président de la République assurait qu'il comprenait la situation des entreprises qui employaient des travailleurs immigrés sans papiers, qu'il était important de maintenir des bas salaires pour garantir la vitalité des petites et moyennes entreprises, et qu'il veillerait à ce que ces entreprises ne soient pas pénalisées.

La vidéo s'était vite répandue dans les réseaux sociaux et les propos du Président avaient été relayés dans les premières heures par deux grands médias, la chaîne d'information TBT et le site lefureteur.com. Le porte-parole de l'Élysée avait immédiatement démenti les propos attribués au Président et avait fait savoir que la vidéo était un faux créé par une agence étrangère qui tentait d'interférer dans les élections françaises. La technique utilisée pour créer la vidéo avait été mise au point par l'entreprise américaine Monkeypaw Productions qui avait tiré parti des algorithmes GAN (*generative adversarial networks*), élaborés par des chercheurs de l'Université de Montréal en 2014.

Contre toute attente, les images créées grâce à l'IA avaient atteint un degré de réalisme stupéfiant en moins de dix ans, si bien qu'une fausse vidéo ne pouvait plus être détectée à l'œil nu.

Ni le démenti de l'Élysée, ni le mea culpa de TBT et du Fureteur, ni encore l'interdiction de diffusion de la vidéo n'avaient eu l'effet espéré. La vidéo était encore consultable sur différents sites étrangers comme le site rassvet.io. Un député du parti populiste de droite en avait profité pour accuser le Président de faire le jeu de l'immigration clandestine et de nuire aux intérêts des Français. Le nombre de gazouillis avec le mot-clic *#Presidentclandestin* avait passé la barre des 300 000 en une semaine. À un mois du premier tour des élections présidentielles, Dominique B., directrice de l'AII, doit présenter un plan pour enrayer les effets dévastateurs de cette fausse information et rétablir les conditions d'une campagne électorale saine. Mais ce matin, le sentiment d'avoir déjà épuisé toutes les solutions l'emporte à l'AII.

ENVIRONNEMENT

La cote environnementale basée sur l'empreinte de carbone

1er février 2025. Pour la cinquième année de suite les températures battent des records de chaleur dans le monde entier. La majorité des pays ayant signé l'Accord de Paris en décembre 2015 n'ont pas tenu leurs engagements en raison des impératifs économiques de court terme, malgré les mises en garde du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC). En conséquence, les villes européennes du C40, le réseau des villes engagées dans la transition écologique, ont accéléré leur coopération pour proposer à leurs habitants un système de permis carbone individuel fortement incitatif, le système ÉcoFit, connecté à leur compte bancaire et aux différentes applications d'achat en ligne : dans ces villes, le prix des biens et services est affiché en euros et en carbone, et chaque citoyen doit viser 4 tonnes d'émission de carbone par an pour l'ensemble de sa consommation. Les personnes qui atteignent cet objectif augmentent leur cote environnementale calculée par l'algorithme ÉcoFit, à partir de leurs données personnelles de consommation. Cette cote leur donne un accès gratuit à de multiples services écoresponsables en transport, éducation, formation et culture.

15 juin 2025. Au moment de passer leur commande de coquilles Saint-Jacques grâce à leur réfrigérateur FrigoMax connecté, Ive et Charles, habitants du 20e arr. à Paris, découvrent ce nouveau système de points auquel ils viennent d'adhérer : Coquilles Saint-Jacques (Provenance : Pérou) : 12 € / 22 kg éq. CO2/kg*[27]. Un message d'avertissement s'affiche : « Cet achat doit rester exceptionnel. Vous ne pourrez pas tenir votre objectif annuel si vous le reproduisez souvent. » Et l'algorithme de recommandation de FrigoMax leur propose alors des coquilles Saint-Jacques de Saint-Brieuc, fraîches, qui coûtent 22,5€ mais seulement 0,25 kg éq. CO2/kg.

15 octobre 2025. Après quelques écarts, et suite aux nombreux messages d'avertissement, Ive et Charles ont fait un effort pour consommer plus sobrement grâce aux recommandations d'ÉcoFit : régime presque végétarien, nouvelle isolation de leur logement, transport en commun et en vélo, contrat d'électricité verte, choix exclusif d'applications avec *data-centers* carbone neutre : c'est qu'au bureau, tout le monde compare maintenant sa cote environnementale !

1er décembre 2025. Grâce à leurs comportements de plus en plus vertueux, Ive et Charles ont réussi à rester juste en-dessous du plafond visé : après 6 mois, ils sont chacun à 1,95 tonne de carbone pour leur consommation globale. De plus, ils ont moins dépensé monétairement, ce qui leur procure une épargne inattendue. Le couple considère alors de réaliser son projet de séjour à Cuba pour Noël et commence à consulter les sites des agences de voyage. Un message leur parvient sur leur téléphone : « Attention à l'*effet rebond* : dépenser vos économies dans un voyage annulerait tous vos efforts ! Pensez à voyager local ! »

[27] kg éq. CO2/kg = kilogramme équivalent carbone ; exprimé ici par kg de produit importé par avion.

MONDE DU TRAVAIL

Forage des données (*data mining*) RH pour optimiser l'ambiance au travail

30 octobre 2025. Pierre-André a enfin décroché un emploi dans un bon bureau d'avocats qui traite notamment du droit de l'environnement, l'un de ses domaines de prédilection.

Après trois semaines de travail, il rencontre Marco aux ressources humaines pour une séance de mentorat personnalisée. Marco fait le point sur l'intégration de Pierre-André, sur ses attentes initiales, ses difficultés, etc. Il lui explique aussi que la firme utilise désormais AmbIA+, une IA d'analyse conversationnelle qui étudie les attitudes des salariés et aide à maintenir une ambiance de travail apaisante et productive. C'est une question d'efficacité. Ainsi, tous les courriels, appels téléphoniques et prises de parole en réunion d'équipe sont analysés pour extraire un historique des humeurs et des émotions des salariés. Ces données sont ensuite rapportées à un laboratoire de recherche en psychologie.

Pierre-André est déstabilisé et même un peu inquiet, mais Marco essaie de le rassurer :

> AmbIA+ fournit une assistance individualisée, elle conseille et entraîne, mais il n'y a pas de sanction. D'ailleurs, AmbIA+ ne mémorise que la forme des interactions, et tous les échanges que vous avez eus jusqu'à présent au bureau se sont bien passés.

Tous, sauf pour le 15 et le 16 octobre derniers. Pierre-André travaillait alors sur le dossier de la nouvelle station d'épuration des eaux usées de la ville de Lille. « Selon AmbIA+, rapporte Marco, vous avez à plusieurs reprises interrompu vos collègues en réunion pour répéter les mêmes idées, ce qui a créé de la tension chez eux. Il faudrait essayer d'exposer vos arguments en une fois, lors du tour de table, pour ne pas perdre de temps. »

Mais ce n'est pas tout :

> Apparemment, l'algorithme a aussi détecté des périodes d'inactivité sur le réseau de plusieurs heures, sans aucun échange avec vos collègues. Ce n'est pas grave en soi, mais c'est mieux de maintenir le contact avec l'équipe. Est-ce que vous vous souvenez de la raison de cette inactivité ?

Pierre-André n'est plus seulement inquiet, il est embarrassé et s'interroge sur la pertinence de ces questions :

> Oui, c'est vrai, j'aime bien travailler avec un crayon sur un rapport papier et je préfère ne pas rédiger directement sur le document collaboratif en ligne… et en effet lors de la réunion du 15, j'apportais une idée nouvelle qui ne me semblait pas bien comprise et je craignais qu'on ne l'oublie. Mais est-ce vraiment un problème ?

Compréhensif, Marco répond qu'il n'y a vraiment aucun problème : « Mais ne vous déconnectez pas de l'équipe, c'est mieux pour la performance collective. Allez, on se revoit dans deux mois. Et bonne chance pour la réunion de demain ! »

ANNEXE 2
Les brèves étudiantes

Simulation 2018, CIÉ-FRQ

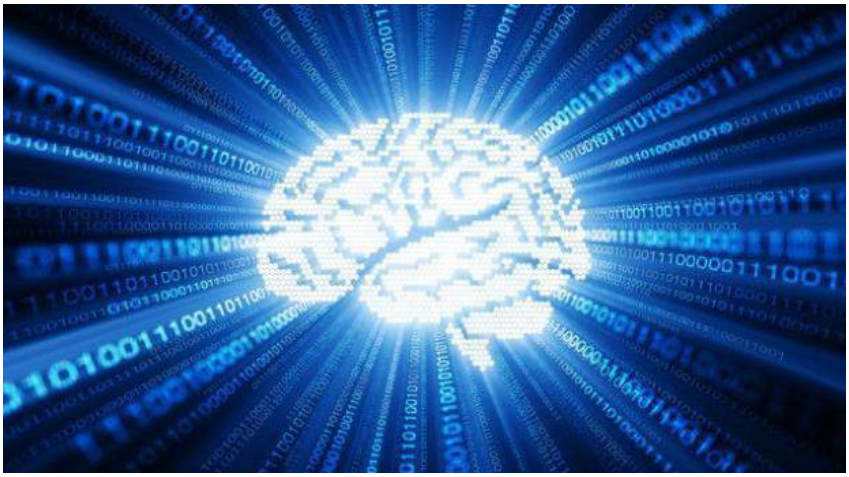
Brève politique sur l'intelligence artificielle

Le présent document est le résultat d'un exercice de simulation, dont l'objectif était d'acquérir des compétences en rédaction et en communication publique. Étant donné le contexte pédagogique dans lequel cette note a été produite, elle n'a pas la vocation, dans les faits, d'être adressée à des décideurs ou à des acteurs de la fonction publique.

La Déclaration de Montréal a choisi de publier ces brèves afin de représenter fidèlement le résultat d'un travail réalisé en 6 heures par les étudiants de la relève et montrer la pertinence d'un tel exercice.

Problématique 1 : Sécurité publique et intégrité des systèmes

Sous-problématique 4 : Les conséquences engendrées par une mauvaise évaluation des capacités de l'intelligence artificielle



Document rédigé par :

Joël Simoneau

Jérôme Gélinas Bélanger

Fidele Ndjoulou

Moumouni Ouiminga

À l'intention du Gouvernement du Canada

Titre de la brève

Pour l'établissement d'un système de responsabilité de l'intelligence artificielle dans les biens de consommation

Cette brève politique expose une démarche qui vise à promouvoir une intelligence artificielle responsable pour la sécurité et la protection des Canadiennes et Canadiens. Elle porte spécifiquement sur les systèmes embarqués. Les recommandations énoncées sont :

1. La mise sur pied d'un organisme pancanadien de certification des systèmes embarqués[1] utilisant l'intelligence artificielle
2. Le développement d'un régime de responsabilité sans faute

De manière concrète, le gouvernement devrait s'atteler dans un premier temps à la création d'un organisme fédéral responsable de la certification obligatoire des systèmes embarqués utilisant l'IA. Dans un deuxième temps, il est impératif de mettre sur place un régime propriétaire sans faute.

Une telle politique permettra d'assurer une meilleure santé et sécurité ainsi qu'une protection légale à tous les Canadiennes et les Canadiens dans leur interaction avec des objets utilisant l'IA, tout en impliquant les entreprises privées dans le processus de la saine utilisation de l'IA.

[1] On qualifie de « système embarqué » un système électronique et informatique autonome dédié à une tâche précise, souvent en temps réel, possédant une taille limitée et ayant une consommation énergétique restreinte. www.futura-sciences.com/tech/definitions/technologie-systeme-embarque-15282/

La société canadienne à l'ère du développement de l'intelligence artificielle

Notre société est en train de vivre une transformation globale basée sur l'évolution du numérique. Autant celui-ci véhicule des informations à une vitesse précédemment inimaginable, qu'il transforme notre rapport avec les objets. Les avancées technologiques récentes en intelligence artificielle (IA) permettent d'imaginer un futur imminent où certaines tâches avec prises de décision redondantes seraient attribuées à des logiciels conçus expressément pour cette fonction. Les véhicules autonomes sont déjà au coin de la rue, les dispositifs médicaux intelligents sont derrière les portes des universités. Une mauvaise médicamentation ou des accidents automobiles sont des dangers qui tendent à être réglés par l'utilisation intelligente et sécuritaire de l'IA, mais il faut aussi s'assurer qu'elle n'en devient pas la cause. Cela représente des inquiétudes énoncées par les Canadiennes et les Canadiens à travers les travaux de la Déclaration de Montréal pour un développement responsable de l'IA.

La régularisation des systèmes embarqués, soit un appareil physique contenant un logiciel utilisant une IA, devrait être un projet d'importance pour le gouvernement canadien. Ceux-ci représentent une implémentation physique et commercialisable d'un produit d'IA, et il serait important d'en assurer une règlementation en amont de leur arrivée prochaine sur le marché canadien. Une prise de décision proactive et l'installation d'un cadre réglementaire permettrait l'encadrement des IA pouvant avoir un impact physique direct sur le peuple canadien.

Ce document propose l'instauration d'un organisme réglementaire de certification des systèmes embarqués utilisant l'IA et d'un régime de responsabilité basé sur le propriétaire sans faute. L'organisme permettrait d'encadrer les normes de sécurité de conception et d'utilisation des systèmes, et le régime permettrait de définir exactement le rapport de responsabilité dans le but de protéger les Canadiennes et les Canadiens, autant légalement qu'au niveau de leur santé et bien-être. La combinaison de ces deux mesures encadrera les systèmes embarqués, de leur commercialisation jusqu'à leur utilisation, ce qui maximisera les impacts positifs du développement de l'IA, en réduisant ses effets néfastes.

Constats et pistes d'action sur le développement de l'intelligence artificielle au Canada

1. Organisme de certification

À l'heure présente, aucun cadre législatif n'existe quant à l'utilisation de l'intelligence artificielle intégrée à des systèmes embarqués au Canada. Ce flou juridique pose un certain nombre de défis pour les différents paliers de gouvernement, notamment le gouvernement fédéral, relativement à leur capacité de structurer la mise en marché et la régulation de ces objets au pays. De façon plus générale, ce manque de structure à ce niveau engendre des complexités juridiques en termes d'évaluation du risque que présentent ces technologies pour le public, mais aussi en termes de l'attribution du poids de la responsabilité advenant un incident découlant de l'utilisation d'une technologie basée sur l'IA.

Face à ces défis, il apparaît nécessaire pour l'État canadien de créer un organisme réglementaire de certification des systèmes embarqués utilisant l'IA, l'office de réglementation nommé l'Agence de Réglementation sur les Biens utilisant l'Intelligence Artificielle (ARBIA), à vocation interdisciplinaire et agissant comme pilier décisionnel. Cet organisme possédera trois principaux axes d'action afin de parvenir à structurer la réglementation de l'IA à l'échelle canadienne: 1) l'investissement dans la recherche et l'innovation permettant le développement de balises législatives basées sur des connaissances techniques, 2) l'instauration de comités experts possédant une bi-spécialisation reposant sur l'IA et leur propre champ d'expertise à l'intérieur des différents ministères pouvant être éventuellement affectés par le développement de l'IA et 3) le développement d'une plateforme réglementaire encadrant la mise en marché et le régime propriétaire sans faute.

L'implication directe du gouvernement canadien dans les cas de problématique de bien utilisant l'IA permettra d'assurer une veille scientifique et sécuritaire proactive, et de protéger légalement les consommateurs canadiens, qui n'auront pas à subir des procès-bâillons.

2. Régime de responsabilité sans faute

On entend par responsabilité l'obligation de répondre d'un dommage devant la justice et d'en assumer les conséquences notamment civiles et pénales envers la victime et/ou la société. Dans un régime de responsabilité sans faute, le gouvernement du Canada sera responsable des accidents physiques ou matériels causés par un bien matériel utilisant l'IA. Dans le cas d'un bien non conforme au processus de certification, le gouvernement canadien peut intenter des actions contre le fabricant.

L'implication directe du gouvernement canadien dans les cas de problématique de bien utilisant l'IA permettra d'assurer une veille scientifique et sécuritaire proactive, et de protéger légalement les consommateurs canadiens, qui n'auront pas à subir des procès-bâillons.

2.1 Secteurs d'activités concernés

L'intelligence artificielle s'applique à plusieurs secteurs d'activité notamment la santé, l'éducation, la sécurité, l'agriculture. Cependant, cette brève politique touche de manière spécifique l'automobile autonome, les dispositifs médicaux et la domotique.

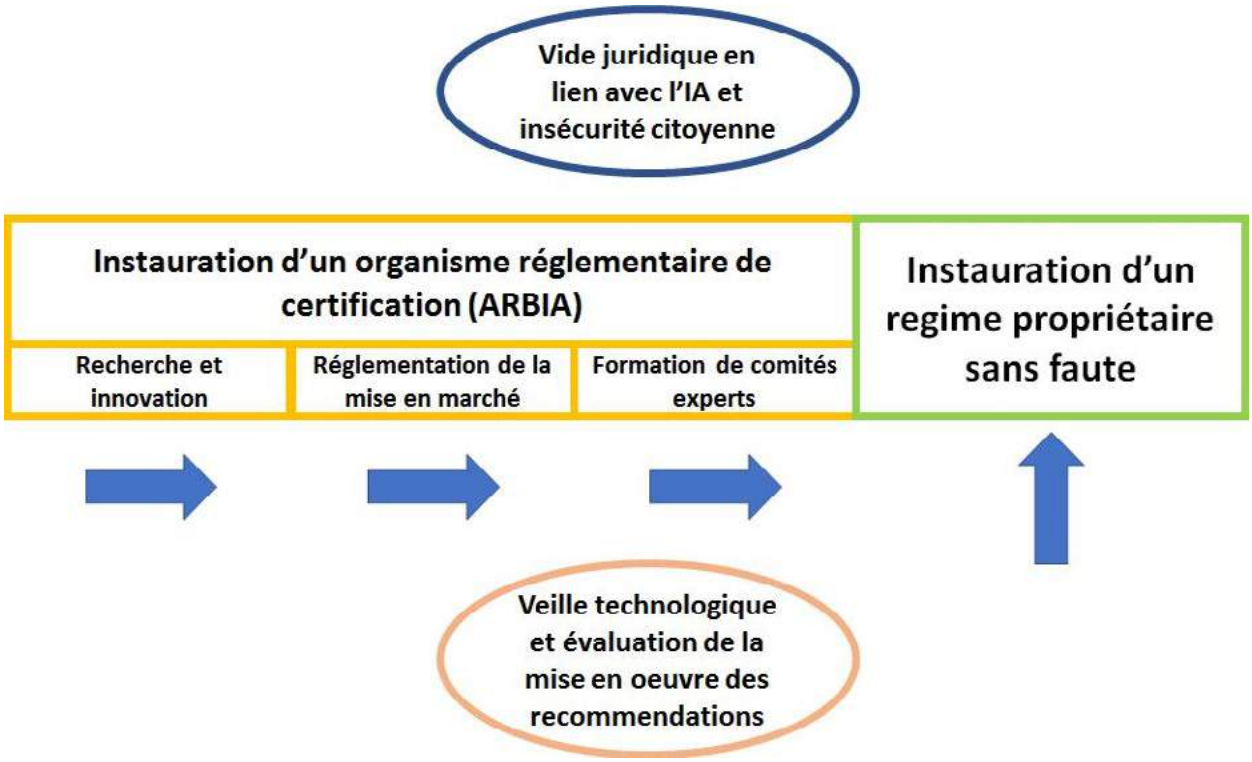
2.1.1 Automobiles autonomes

Avec le développement de l'intelligence artificielle, le secteur de l'automobile a connu une transformation radicale. Une nouvelle catégorie d'automobile dite automobile autonome est mise sur le marché. Néanmoins, ces voitures ont déjà causé des accidents aux États-Unis, par exemple l'accident mortel causé par une voiture autonome en Floride en mars 2018. Compte tenu de l'absence des règles spécifiques à la circulation des automobiles autonomes, et dans le souci d'apporter une meilleure protection aux citoyens, il est impératif de définir un cadre réglementaire au niveau fédéral. Ce cadre va fixer la responsabilité des parties prenantes, à savoir l'État et les compagnies propriétaires des voitures autonomes. Une faute liée aux défaillances est une faute de la compagnie responsable et propriétaire de la voiture, tandis qu'une utilisation faite par l'individu utilisateur est une faute de sa part.

2.1.2 Dispositifs médicaux

Pour faciliter la vie des personnes vivant avec le diabète, une pompe à insuline a été développée. C'est un système intégré composé de trois parties (boîtier,

composantes électroniques, cathéter) qui libère automatiquement de l'insuline. Le fonctionnement de ce système nécessite impérativement une formation du patient, une autosurveillance glycémique et un suivi médical rapproché. Cela engendrera une responsabilité du patient en cas de mauvaise utilisation de sa part. Les conséquences pourraient être énormes au point d'engendrer d'éventuels cas de décès.



Retombés et recommandations

Dans l'objectif d'assurer la santé et la sécurité des Canadiennes et Canadiens face aux systèmes embarqués utilisant l'IA et de maximiser les impacts positifs du développement de l'IA, nous émettons les recommandations suivantes.

Recommandation 1

Création de l'Agence de Réglementation sur les Biens utilisant l'Intelligence Artificielle (ARBIA)

Retombées :

- Coût nul pour le gouvernement canadien

  Le financement de l'organisme de certification et des actions légales sera couvert par une licence de fabrication des systèmes.
- Fiabilité des systèmes d'IA pour la santé et sécurité des Canadiennes et Canadiens.

  Par le respect de normes définies par des comités experts, normes qui seront mises à jour selon les cas vécus.

Recommandation 2

Instauration d'un régime propriétaire sans faute

Retombées :

- Accessibilité judiciaire améliorée dans le cas de faute des fabricants.

  Le système sans faute donne la responsabilité de la poursuite judiciaire au gouvernement canadien, qui a plus de ressources que les citoyennes et citoyens individuellement.
- Promotion de l'implication sociale des entreprises.

  Considérant leur responsabilité directement impliquée, les fabricants vont être encouragés à développer des mécanismes d'utilisation sécuritaire de leurs produits.
- Veille technologique et sécuritaire du gouvernement canadien

  Considérant l'implication directe du gouvernement canadien dans les processus judiciaires, celui-ci assure une veille permanente dans la gestion saine des IA.

Bibliographie :

[1] Pour aller plus loin voir Palmer, Vernon. « Trois principes de la responsabilité sans faute » (1987) 39:4 Revue internationale de droit comparé; Mémeteau, Gérard. « Un point sur la responsabilité civile du fait des prothèses » (2013) 2013:123 Médecine & Droit 175-180; Jacob, Julien. « Prévention des risques technologiques à l'aide de la responsabilité civile en présence d'une innovation à double impact » (2013) 202:1 Économie & amp; prévision 1-18.
[2] https://diabetnutrition.ch/les-traitements/la-pompe-a-insuline-quest-ce-que-cest/
3https://ici.radio-canada.ca/info/videos/media-7560667/premier-accident-mortel-impliquant-une-voiture-autonome source consultée le 18 octobre 2018

Simulation 2018 dans le cadre des J2R

« Politique et intelligence artificielle »

Le présent document est le résultat d'un exercice de simulation, dont l'objectif était d'acquérir des compétences en rédaction et en communication publique. Étant donné le contexte pédagogique dans lequel elle a été produite cette note, elle n'a pas la vocation, dans les faits, d'être adressée à des décideurs ou à des acteurs de la fonction publique.

La Déclaration de Montréal a choisi de publier ces brèves afin de représenter fidèlement le résultat d'un travail réalisé en 6 heures par les étudiants de la relève et montrer la pertinence d'un tel exercice.

Fausses nouvelles, vrais enjeux : s'éduquer pour y faire face

Présenté aux membres du jury

Par

Jean Clairemond César, étudiant au doctorat en éducation, Université de Sherbrooke
Isabelle Dufour, inf., candidate au doctorat, Université de Sherbrooke
Gaël Grissonnanche, post-doctorant en physique, Université de Sherbrooke
Philippe Lebel, doctorant en microbiologie, Université de Montréal

18 octobre 2018

Tables des matières

But de la brève

Émettre des recommandations auprès d'un décideur public en vue d'offrir une ou plusieurs pistes de solution à un problème spécifique découlant d'une des trois problématiques décrites dans le document élaboré par l'équipe de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle. Une problématique sera attribuée par équipe et les participants devront identifier les éléments suivants :

- Le **problème**
- La ou les **solution**(s) recommandée(s) et les **répercussions** sur la population visée et non visée
- Le **décideur public** impliqué
- Les **facteurs environnementaux** pouvant faire obstacle à la mise en œuvre de la ou des solution(s) recommandée(s).

Couverture (1 page)

La première page présente une synthèse de la brève politique. Elle présente la pertinence de la brève et ses grandes lignes, les conclusions clés et la marche à suivre.

Cette brève politique présente le problème des fausses nouvelles sur l'internet. Aujourd'hui la proportion de Canadiens qui consomme de l'information en ligne a dépassé celle des médias traditionnels. L'efficience de cette technologie repose sur l'intelligence artificielle (IA) en offrant des contenus filtrés selon le comportement et l'intérêt de l'utilisateur. De nos jours, plusieurs acteurs sociopolitiques ont levé le drapeau rouge sur cet enjeu de société. D'un autre côté, les grandes entreprises de médias sociaux telles que Google, Facebook, Amazon et tant d'autres proposent déjà des mesures pour limiter le potentiel propagandiste de leur algorithme, et la définition d'une fausse nouvelle ne fait pas consensus. La pertinence de notre brève se situe dans la nécessité d'augmenter l'esprit critique au sein de la population québécoise. L'absence d'esprit critique peut occasionner plusieurs problèmes en éducation et en santé et dans d'autres domaines. Destiné aux ministres concernés par la stratégie numérique, ce document présente plusieurs recommandations sur l'importance de l'éducation aux médias et à l'information au Québec.

Introduction (1 page)

Cette section décrit l'objectif principal de la brève et le problème politique. Elle établit un lien entre les données probantes et le problème.

L'avènement de l'internet apporte aux citoyens la démocratisation de l'accès à l'information à travers des moteurs de recherches intelligents et des médias sociaux. Aujourd'hui, la proportion de Canadiens consommant de l'information en ligne a dépassé celle des médias traditionnels. L'efficience de cette technologie repose sur l'intelligence artificielle (IA) en offrant des contenus filtrés selon le comportement et l'intérêt de l'utilisateur. Tandis que certaines études montrent que cette exposition partiale à l'information tend à engendrer chez l'utilisateur une confirmation systématique de sa pensée, d'autres en revanche arguent que celui-ci n'a jamais été exposé à une telle diversité de sources lorsque comparé à la presse écrite, à la télévision, à la radio, etc.

C'est dans ce contexte qu'émerge sur la scène internationale la notion de fausse nouvelle comme un enjeu de désinformation massive dans une société démocratique. L'usage d'IA comme en a fait la firme Cambridge Analytica aux États-Unis a montré au monde le niveau de déstabilisation sociétale que cette technologie peut engendrer. Alors que les grandes entreprises de médias sociaux telles que Google, Facebook, Amazon et tant d'autres proposent déjà aujourd'hui des mesures pour limiter le potentiel propagandiste de leur algorithme, la définition d'une fausse nouvelle ne fait pas consensus. En effet, selon le Global News, près de 58% des Canadiens définissent celle-ci comme une histoire pour laquelle les faits sont faux. Cependant, 46% l'emploient pour désigner les nouvelles de journaux et les discours de personnalités politiques n'exprimant qu'un unique côté des faits. Encore, ce même chiffre désigne le pourcentage pensant que ce terme est uniquement utilisé par les politiciens pour discréditer les médias qui les critiquent. À l'autre bout du spectre, des actions pour valoriser l'esprit critique de l'utilisateur demeurent une avenue qui doit être envisagée.

La problématique amenée par l'essor des fausses nouvelles dans les médias est importante et est susceptible d'avoir un impact important sur la population. À cet égard, l'objectif de cette brève est d'augmenter la sécurité de la population et leur éducation face aux fausses nouvelles.

Données probantes et analyse (3 pages)

Cette section représente le cœur de la brève politique. La qualité de cette section est jugée par la pertinence des données présentées, des interprétations tirées de ces données, ainsi que de leurs apports et de leurs limites. Elle peut contenir des graphiques, des tableaux et des schémas.

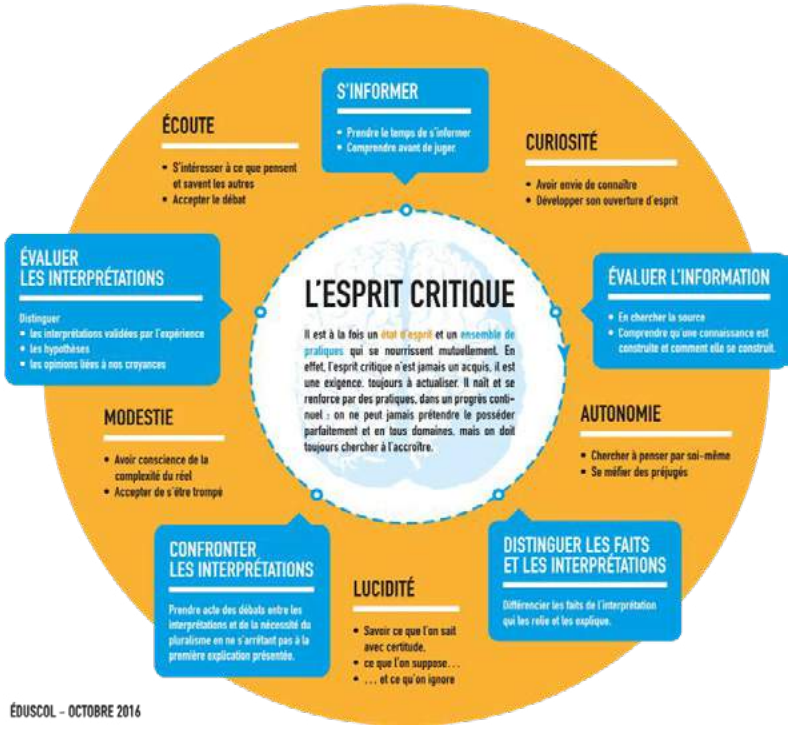
Selon Jeff Yates, expert québécois de la question, une fausse nouvelle se définit comme « une information soit carrément fausse, détournée, exagérée ou dénaturée à un point tel qu'elle n'est plus véridique, et présentée comme une vraie nouvelle dans le but de tromper les gens. Cela peut être fait pour générer des clics et des partages sur les réseaux sociaux, pour atteindre des objectifs quelconques (politiques, idéologiques, économiques, etc.) ou simplement pour se moquer de la crédulité des lecteurs ». Sujet de débats socio-économiques, les fausses nouvelles dans les médias ont connu un essor marqué durant les dernières années, et principalement avec le développement de l'IA. En effet, des méthodes associées à l'IA sont utilisées par les sites de médias sociaux et peuvent procéder de façon automatique à la diffusion de fausses nouvelles.

En Amérique du Nord, environ 60% de la population croit que la dispersion de telles informations dans les médias cause de la confusion. Les enfants et les adolescents sont particulièrement à risque d'attribuer du crédit et de participer à leur diffusion. Le manque de consensus dans la définition d'une fausse nouvelle contribue à l'incertitude vécue par la population, 45% des Canadiens en ayant une vision erronée. À l'ère numérique, la nécessité de développer une pensée critique concernant les informations transmises par les médias se positionne donc comme un enjeu central.

Selon Vallerand, la pensée critique est une pensée responsable qui s'appuie sur des critères et qui est sensible au contexte et aux autres (Vallerand, 2016). L'IA pourrait être utilisée pour développer l'esprit critique des jeunes et les former au doute constructif. Les jeunes du Québec apprendront comment mettre en perspective une information diffusée sur le web. En ce sens, les décideurs politiques donneront les moyens nécessaires pour y arriver.

En France par exemple, le développement de l'esprit critique est au centre de la mission assignée au système éducatif français, comme le présente le modèle de l'esprit critique d'Eduscol. Il est renforcé par l'attention désormais portée à l'éducation aux médias et à l'information. Le travail de formation des élèves au décryptage du réel et à la construction, progressive, d'un esprit éclairé, autonome et critique est essentiel. L'esprit critique est une compétence essentielle du citoyen et de la citoyenne du 21^{e} siècle. Analyser une source, mettre en perspective une image ou une information, en extraire l'essentiel, critiquer le contenu, se questionner sont autant de savoirs numériques nécessaires à l'exercice d'une citoyenneté avisée. Une étude faite en 2017 au Royaume-Uni a montré que seulement 4 % de la population testée avait été capable d'identifier correctement les vraies des fausses nouvelles. Ce résultat est inquiétant, notamment en termes de sécurité publique. Pensons par exemple au mouvement anti-vaccination qui

cause un retour en force de maladies mortelles, tel que la coqueluche aux États-Unis, malgré qu'il a été démontré depuis longtemps que les causes de ce mouvement sont fausses.



Selon les experts de la pensée critique, Christopher DiCarlo et l'auteur du *Petit cours d'autodéfense intellectuel*, Normand Baillargeon, la solution passe par l'éducation. En effet, il est préférable que l'école forme une jeunesse plus critique de ce qu'elle consulte plutôt que de faire confiance aux grandes entreprises privées du web pour autocensurer leur contenu.

D'ailleurs on retrouve plusieurs initiatives, ailleurs comme chez nous, qui vont dans ce sens. En France, plus précisément en Haute-Savoie, des enseignantes ont créé une habitude locale où elles prennent une heure par semaine pour sensibiliser leurs élèves à détecter les fausses nouvelles. Cette initiative, accueillie avec enthousiasme par les élèves, semble rapidement porter fruit puisque ces jeunes de 10 ans ont déjà développé les réflexes de vérifier d'où proviennent des images-chocs qui publicisent de fausses nouvelles sensationnalistes, par exemple.

Plus près de chez nous, depuis mai 2018, un nouveau programme s'implante dans les écoles ontariennes : Actufuté. Ce programme se veut une collaboration entre la Fondation pour le journalisme canadien et l'organisme CIVIX qui est responsable du programme Vote étudiant. Ce dernier prend vie autour des périodes d'élections et encourage la participation citoyenne des 9 à 19 ans. C'est dans ces périodes riches en nouvelles qu'Actufuté viendra aider les élèves à démystifier le vrai du faux.
Au Québec, « École branchée », un organisme sans but lucratif (OSBL) propose des outils aux enseignantes et aux enseignants pour intégrer ces considérations dans leur programme de tous les jours. Malheureusement, à ce jour, seulement 15 % à 20 % du corps enseignant est rejoint par l'organisme. Démonstration qu'une intervention gouvernementale est nécessaire pour offrir une protection équitable à tous nos jeunes contre ce fléau. Ce faisant, la jeunesse pourra aussi transmettre cette information et conscientiser ses proches à la problématique.

Pour y arriver, le Plan d'action numérique en éducation et en enseignement supérieur, annoncé à l'été 2018, prévoit quelque 900 millions de dollars pour, justement, préparer la génération de demain à ce nouvel environnement numérique. Le gouvernement du Québec pourrait ainsi soutenir les services d'« École branchée », voire même intégrer son contenu au cursus normal de l'éducation primaire et secondaire.

L'intégration de cette notion d'éducation directement au cursus scolaire vient contrer l'obstacle environnemental principal. Le désir des professeurs d'assurer le développement de leurs étudiants pourra aussi agir à titre de facilitateurs.

Répercussions sur les politiques et recommandations (1 page)

Cette section présente les recommandations proposées et les répercussions anticipées. Ces recommandations et ces répercussions peuvent s'organiser autour de thèmes, de parties intéressées ou d'échéancier.

L'argumentaire soulevé met en lumière plusieurs défis soulevés par l'IA. Entre autres, elle contribue à la diffusion de masse de fausses nouvelles. Cela cause de la confusion au sein de la population, une perte de confiance envers les sources d'information. Les jeunes et les adolescents sont particulièrement sensibles aux fausses nouvelles, leur capacité de raisonnement et leur esprit critique étant en construction. L'implication des instances gouvernementales est donc primordiale pour assurer la protection et l'éducation des populations, et particulièrement des jeunes, sur la problématique des fausses nouvelles. Les recommandations adressées font appel à des notions de vigilance et d'éducation.

Vigilance

Notre recommandation :

- Alerter la population québécoise sur la dissémination de fausses nouvelles

L'objectif étant de favoriser le développement et l'utilisation d'IA spécialisée pour détecter les fausses nouvelles diffusées sur les médias sociaux.
Cette initiative s'inscrit également en parallèle avec le Détecteur de rumeurs, où les alertes du CVMQ, en cas de détection d'une fausse nouvelle de grande importance, pourront être diffusées.
La création du Comité de vigilance numérique du Québec sera annexée à l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique.

Éducation

Notre recommandation :

- Offrir des outils au corps enseignant québécois pour intégrer la conscientisation face aux fausses nouvelles dans l'éducation, dès l'école primaire.

Cette mesure aura deux buts : préparer directement cette génération à affronter le fléau des fausses nouvelles et les inciter à répandre ces bonnes pratiques auprès de leurs proches.

Grâce au financement déjà prévu pour le Plan d'action numérique en éducation et en enseignement supérieur ainsi qu'aux initiatives déjà en place, il sera possible de protéger la population québécoise sans investissement supplémentaire et sans réinventer la roue. À court terme, la promotion de ces outils auprès des enseignantes et des enseignants aura déjà un impact et il sera possible de penser intégrer ces enseignements au cursus normal à moyen terme.

Tableau 1. Grille d'évaluation des brèves politiques

Critères	Tous les points	- 1 p	- 2 p	- 3 p
Couverture	La synthèse présente la pertinence de la brève et ses grandes lignes, les conclusions clefs et la marche à suivre.	Des éléments sont manquants	La synthèse est manquante	
Introduction	Cette section décrit très bien l'objectif principal de la brève et le problème politique. Elle établit un lien entre les données probantes et le problème.	Cette section décrit bien l'objectif principal de la brève et le problème politique. Elle établit un lien entre les données probantes et le problème.	Cette section décrit convenablement l'objectif principal de la brève et le problème politique. Elle établit un lien entre les données probantes et le problème.	Cette section est absente
Données probantes et analyse	Cette section est très pertinente au regard du problème; Les interprétations sont justes et convaincantes; Les facteurs environnementaux (socio-politico-économico-culturels) sont très bien pris en compte dans la possible intégration des recommandations; Les apports et les limites sont très bien identifiés.	Cette section est pertinente au regard du problème; Les interprétations sont justes; Les facteurs environnementaux (socio-politico-économico-culturels) sont bien pris en compte dans la possible intégration des recommandations; Les apports et les limites sont bien identifiés.	Cette section est plutôt pertinente au regard du problème; Les interprétations sont plutôt justes; Les facteurs environnementaux (socio-politico-économico-culturels) sont pris en compte dans la possible intégration des recommandations; Les apports et les limites sont identifiés.	Cette section n'est pas pertinente au regard du problème; Les interprétations sont erronées; Les facteurs environnementaux (socio-politico-économico-culturels) ne sont pas pris en compte dans la possible intégration des recommandations; Les apports et les limites ne sont pas correctement identifiés.
Répercussions et recommandations	Les recommandations proposées sont très pertinentes et les répercussions anticipées très bien identifiées.	Les recommandations proposées sont pertinentes et les répercussions anticipées sont bien identifiées.	Les recommandations proposées sont plus ou moins pertinentes et les répercussions anticipées plus ou moins bien identifiées.	Les recommandations proposées ne sont pas pertinentes et les répercussions anticipées ne sont pas bien identifiées.
Qualité de la présentation orale	La présentation de la brève est très convaincante.	La présentation de la brève est convaincante.	La présentation de la brève est peu convaincante.	La présentation de la brève n'est pas convaincante.
Total des points	/15			

Simulation 2018 dans le cadre des J2R

« Politique et intelligence artificielle »

Le présent document est le résultat d'un exercice de simulation, dont l'objectif était d'acquérir des compétences en rédaction et en communication publique. Étant donné le contexte pédagogique dans lequel a été produite cette note, elle n'a pas la vocation, dans les faits, d'être adressée à des décideurs ou à des acteurs de la fonction publique.

La Déclaration de Montréal a choisi de publier ces brèves afin de représenter fidèlement le résultat d'un travail réalisé en 6 heures par les étudiants de la relève et montrer la pertinence d'un tel exercice.

Gouvernance publique, privée ou participative : les communs numériques

Présenté aux membres du jury

Par

Thomas Bousquet
Alexandre Côté, PhD(c)
Christian Kouakou, PhD(c)

18 octobre 2018

Tables des matières

Couverture

Vue d'ensemble

L'évolution rapide de la technologie et la science entourant l'intelligence artificielle (IA) exposent certaines brèches dans les politiques et les règlementations actuelles qui concernent ce secteur de développement. Le gouvernement doit se pencher sur ce problème qui soulève de vives inquiétudes pour la population qui s'interroge sur la protection de sa vie privée. L'inquiétude reste présente du côté des entreprises privées qui elles, ont incessamment besoin d'alimenter leur système d'IA avec des données de plus en plus complexes et précises. Le défi du gouvernement est de trouver une forme de gouvernance de l'IA qui répond au mieux aux besoins des différents acteurs concernés.

L'objectif principal de cette brève politique proposée par notre Groupe de travail ponctuel sur l'utilisation de l'IA et des données communes est donc de *fournir une méthode de travail et de réflexion* afin de répondre adéquatement à ces interrogations, en s'assurant de considérer les intérêts distincts de la population et du secteur privé.

Intérêts des acteurs impliqués

Population	- Protéger ses données personnelles - Être rassuré par l'indépendance des instances faisant usage de ses données
Gouvernement	- Assurer la protection du public - Stimuler la croissance économique du secteur technologique
Privé	- Connaître une croissance économique stable - Développer et améliorer les connaissances concernant l'IA

Problèmes soulevés par ces intérêts distincts :

- Politique actuelle mal adaptée
- Inquiétudes des citoyens
- Flou dans les responsabilités lors d'incidents
- Absence de méthodes pour gérer les problèmes liés à l'IA

Plusieurs pistes de solutions

- Création d'un organisme de régulation indépendant
- Favorisation d'une responsabilité partagée relativement aux données communes
- Ajustement des lois et réglementations en vigueur afin de les adapter aux nouvelles réalités technologiques

Introduction

L'évolution rapide de la technologie et la science entourant l'intelligence artificielle (IA) exposent certaines brèches dans les politiques et les règlementations actuelles qui concernent ce secteur de développement. Bien que certaines lois soient déjà en place, la vitesse de l'appareil public peut difficilement rattraper celle de la croissance technologique, et les règles en place deviennent rapidement inadaptées.

Les inquiétudes de la population

Cette inadéquation des politiques publiques en matière d'encadrement de l'IA et de l'utilisation des données servant à sa croissance inquiète la population québécoise. Une consultation récente initiée par un groupe d'experts composé, entre autres, de gens de l'Université de Montréal, de l'Université McGill et de l'Institut de valorisation des données (IVADO), a permis d'identifier certaines préoccupations clés des citoyens vis-à-vis les enjeux actuels concernant notamment :

- La responsabilité face aux données et à l'IA ;
- La protection de la vie privée des individus ;
- La valeur marchande des données partagées ;
- Les risques de mise en place d'un monopole, et de conflits d'intérêts entre les différents acteurs touchés ;
- Ainsi que l'indépendance des différents acteurs qui interviennent dans le domaine.

Les considérations face au secteur privé

La croissance économique québécoise étant de plus en plus liée aux nouvelles technologies, à l'exploitation des données et au développement de l'IA, il est important pour le gouvernement – malgré les inquiétudes soulevées – de ne pas laisser le secteur privé au dépourvu. L'accès aux données de la population est le carburant de cet important moteur économique qui doit manifestement être régulé, mais pour qui une marge de manœuvre doit être maintenue.

Le défi de la gouvernance

Les trois acteurs généraux qui sont touchés par la problématique – le gouvernement, la population et le secteur privé – ressentent déjà les impacts du manque d'ajustement des politiques actuelles. L'exemple récent des piratages de données des grands services technologiques comme Facebook et Google – utilisés par des centaines de milliers de Québécois – expose bien ce phénomène : la population est ultimement la victime, blâmant à la fois le secteur privé et le gouvernement pour les dommages encourus.

Dans cette optique, il est donc primordial pour le député délégué à la transformation numérique gouvernementale de se positionner et de répondre aux questions et inquiétudes soulevées par la société civile et le secteur privé :

1. Qui est responsable face aux données communes ?
2. Quel appareil assure la protection du public ? Fonctionne-t-il adéquatement ?
3. Quel niveau de transparence est optimal ?
4. Comment peut-on stimuler la croissance économique dans le secteur des nouvelles technologies, tout en maintenant la confiance de la population face au processus ?

L'objectif principal de cette brève politique proposée par notre Groupe de travail ponctuel sur l'utilisation de l'IA et des données communes est donc de *fournir une méthode de travail et de réflexion* afin de répondre adéquatement à ces interrogations, en s'assurant de considérer les intérêts distincts de la population et du secteur privé.

Données probantes et analyse

Le problème de l'immigration discriminante

Le gouvernement canadien serait en expérimentation de l'utilisation de l'IA pour le tri des demandes de visa et d'immigration. Cette information a été publiée dans un rapport du Citizen Lab et relayée dans les médias canadiens.

L'une des auteurs du rapport a souligné que « sans garanties et mécanismes de surveillance appropriés, utiliser l'IA pour déterminer l'immigration et le statut de réfugié est très risqué ». Il est clair que l'utilisation de l'IA pourrait aider grandement à accélérer le tri et traitement de données d'immigration. Cela ne saurait cependant et en aucun cas outrepasser la décision discrétionnaire liée au droit à l'immigration qui ne saurait être laissé à une machine ou un algorithme ; algorithme qui plus est, n'est pas sans biais car fortement dépendant des considérations des personnes qui programment. Un arbitrage est à faire entre « rapidité dans le traitement des demandes » et « sélection discrétionnaire des dossiers ».

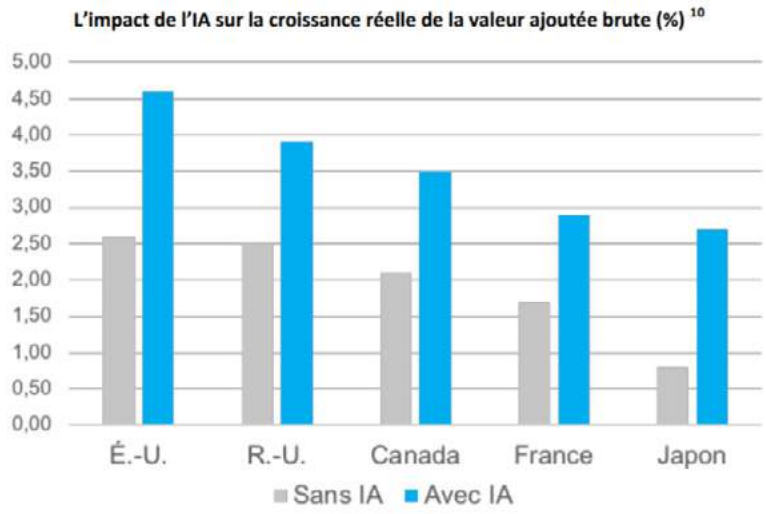
En outre, le rapport à l'immigration de l'équipe responsable de l'algorithme pourrait fortement déteindre sur le résultat final de sélection. D'une part, l'on pourrait avoir un processus de sélection moins rigoureux, voire laxiste, qui laisserait entrer sur le territoire québécois des personnes ne remplissant pas les conditions requises pour l'immigration. D'autre part, un processus plus strict pourrait ôter la possibilité aux personnes remplissant les conditions d'y accéder. Car rappelons-le, la sélection initiale aura été faite non pas par choix discrétionnaire, mais par une machine intelligente.

La confidentialité des données

Les plus gros acteurs en matière de la gestion des données d'utilisateurs, dont font partie Google et Facebook, mettent en place des actions correctives dans leur gestion des données personnelles : ces données ont une valeur monétaire importante et ne doivent être utilisées que dans le cadre où elles ont été fournies par l'utilisateur du service. Les piratages sont récurrents comme l'ont montré les évènements récents (500 000 comptes Google et 29 millions d'utilisateurs Facebook piratés) et la population est donc inquiète de l'utilisation qui peut être faite de ses données.

Les *Big Data* ont également un intérêt dans la santé des populations : des intelligences artificielles sont en développement pour aider au diagnostic médical et nécessitent donc des données sensibles, très personnelles et donc qui font partie d'un haut niveau de confidentialité. La société québécoise a donc évoqué ses inquiétudes quant à la gestion de ces données lors de l'étude qui a été menée sur une partie de la population lors de la rédaction de la Déclaration de Montréal pour l'IA responsable.

Les données personnelles sont donc classifiées selon leur confidentialité et sont donc disponibles à plusieurs niveaux. L'utilisateur sait à qui et dans quel but chaque organisme collecte des données à son sujet.
Il devient donc important qu'un comité multisectoriel neutre se charge de veiller à ce que les possesseurs de ces grandes quantités de données les utilisent de manière éthique pour éviter que des centaines de milliers d'utilisateurs québécois de ces services voient leurs données diffusées sur le web.



De nombreux pays font du développement de l'IA une priorité majeure en investissant dans des organismes de recherche visant à progresser dans ce domaine, comme le montre ce graphique issu de la communication de « Économie, science et innovation Québec » à propos de l'essor de l'écosystème québécois en intelligence artificielle publié en mai 2018.

Le Québec ressent également ce besoin et prévoit des fonds à la recherche dans ce domaine, c'est pourquoi cette volonté doit alors s'étendre à la protection des populations et de leurs données sans pour autant empêcher les acteurs privés et universitaires de progresser dans leurs recherches et leurs innovations.
D'après le projet de mettre le numérique au service du bien commun au Québec (disponible sur le site économie.gouv.qc.ca) d'ici 5 ans, le Québec prévoit mettre à la disposition de la population une transformation numérique des municipalités qui se traduit par une collecte de données continue.

De plus, le Québec prévoit également que les citoyens pourront interagir de façon numérique avec les services de santé et sociaux d'ici les prochaines années. Toutes les données nécessaires à ces services nécessitent des données sensibles sur les citoyens et il est donc très important pour la province de se mobiliser pour protéger les utilisateurs contre une mauvaise utilisation de ces données dans tous les domaines confondus, que ce soit les services publics ou bien les entreprises privées de services.

Le développement du numérique, prisé par le Québec, doit donc faire évoluer les règlements relatifs à la protection des données des citoyens en faisant travailler les entreprises leaders de l'IA conjointement avec les pouvoirs publics et les intérêts des utilisateurs.

Répercussions sur les politiques et recommandations

Nous constatons au final un manque d'adaptation des politiques actuelles face à l'IA et la gestion des données qui alimentent sa croissance. Cette problématique engendre un flou qui touche non seulement le gouvernement et la classe politique, mais également la population et le secteur privé.

Il existe actuellement un manque de clarté quant à la responsabilité de ces trois acteurs face aux données communes (1). Non seulement la *Loi sur la protection des renseignements personnels et des documents électroniques* (LPRPDE ; fédérale) et la *Loi sur la protection des renseignements personnels dans le secteur privé* (provinciale) ne répondent pas à ce manque, elles ne se sont pas non plus adaptées assez rapidement aux nouvelles réalités technologiques, ce qui risque de créer un doute dans la population quant à sa protection et la protection de ses données personnelles (2). Un certain degré de transparence (3) est évidemment requis afin de pallier cet aspect de la problématique, tout en gardant à l'esprit l'importance de ne pas entraver le développement économique du secteur technologique au Québec.

Recommandations

Dans cette optique :

1. Nous emboîtons le pas du *Citizen Lab* de l'Université de Toronto et recommandons la création d'un organisme indépendant de régulation de l'utilisation des données communes. À la différence des chercheurs torontois, nous recommandons cependant que cet organisme soit de juridiction provinciale afin de prendre en considération les particularités de la population québécoise.

2. Nous favorisons une responsabilité partagée des données communes, alimentée par ce nouvel organisme. Ce dernier serait chargé, entre autres, de l'éducation de la population en matière de protection des données personnelles et de la surveillance du secteur privé quant à l'utilisation de ces données.

3. La création de cet organisme viendrait également répondre à la deuxième question soulevée par notre analyse de la situation, soit l'identité de l'appareil veillant à la protection du public. Il serait maintenant clair aux yeux de la population qu'une entité veille à ses intérêts, entre autres en s'assurant – à travers des recommandations émises à l'endroit du gouvernement – de l'adéquation des lois et réglementations concernant l'IA et la gestion des données.

4. Nous suggérons fortement que le nouvel organisme s'assure qu'un certain niveau de transparence soit respecté, autant par le gouvernement que par le secteur privé. Le type de données utilisées, leurs sources, les buts de leur utilisation et leur portée devraient être de nature publique.

5. Nous recommandons qu'un volet économique soit intégré à la mission de l'organisme afin que celui-ci soit constamment en phase avec le marché, s'assurant que les politiques mises en place permettent d'atteindre le parfait équilibre entre croissance et respect du milieu.

6. Finalement, nous recommandons que le nouvel organisme développe une méthode de réflexion permettant d'adapter les lois et réglementations concernant l'IA et les données communes aux changements rapides et fréquents inhérents au secteur des hautes technologies.

< >
Déclaration de Montréal
IA responsable_
</ >

PARTIE 5

RAPPORT DE LA COCONSTRUCTION EN LIGNE ET DES MÉMOIRES REÇUS

TABLE DES MATIÈRES

RÉDACTION

MARTIN GIBERT, conseiller en éthique pour IVADO et chercheur au Centre de recherche en éthique

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

1. INTRODUCTION

En novembre 2017, au terme d'un congrès organisé par l'Université de Montréal au Palais des congrès de Montréal, était lancée la première étape de la *Déclaration de Montréal pour un développement responsable de l'IA*. Une version préliminaire de cette Déclaration articulée autour de 7 principes allait servir de base à une phase de coconstruction débouchant sur une nouvelle version. Si les ateliers de discussion ont permis de consulter les citoyens et experts, d'autres moyens de participer à la réflexion collective étaient possibles : 1) en répondant à un questionnaire en ligne accessible sur le site de la Déclaration (www.declarationmontreal-iaresponsable.com), et 2) en envoyant des mémoires sur un ou plusieurs aspects de la Déclaration. Le présent rapport propose la synthèse des mémoires reçus et des réponses au questionnaire. Le rapport sur les ateliers de coconstruction est également disponible sur le site web de la Déclaration.

2. LE QUESTIONNAIRE EN LIGNE

Le questionnaire en ligne était composé de 35 questions, cinq pour chaque principe. Il a retenu l'attention de 83 répondants, dont 17 anglophones. Comme on pourra le voir dans la synthèse, plusieurs ont une connaissance avancée de l'IA et des enjeux éthiques et sociaux que soulève son développement.

Les questions sont présentées dans l'ordre du questionnaire qui reprend le plan de la Déclaration préliminaire. La Déclaration révisée étant plus complète (elle se compose de dix principes), les nouveaux principes pertinents ont été ajoutés entre parenthèses à ceux de la version préliminaire.

BIEN-ÊTRE (ENVIRONNEMENT, PRUDENCE)

1. COMMENT L'IA PEUT-ELLE CONTRIBUER AU BIEN-ÊTRE ?

Une question générale qui suscite beaucoup de réponses, et des réponses très variées. On invoque beaucoup les espoirs pour la santé et l'aide aux personnes âgées ou en situation de handicap. L'IA semble aussi un espoir pour diminuer les impacts environnementaux même si on note que « le développement de l'IA a une empreinte écologique (donc un coût direct sur le bien-être) souvent négligée, bien que non négligeable ». Plusieurs notent que l'IA pourrait remplacer l'humain dans des tâches à risque. L'aspect « aide à la prise de décision » est aussi mentionné plusieurs fois, en particulier sous la forme d'un possible assistant personnel qui pourrait aussi nous assister dans nos recherches d'information.

On s'attend aussi à ce que l'IA améliore la productivité et nous libère des tâches répétitives et routinières. Elle pourrait aussi anticiper nos attentes et nos besoins ou simplement passer l'aspirateur à notre place. Une disposition importante : l'IA améliorera notre bien-être « à condition que nous vivions dans une véritable démocratie, qu'elle soit au service de tous et pas seulement au service de quelques privilégiés ».

EXTRAITS CHOISIS

"The AI or any technology will create a lot more value for the rural population than the urban population. A single smartphone can provide immense value, and anything that can collect data is a breeding ground for AI: Better education, better farming technology (Eg: crop analysis, robot-farming)."

2. EST-IL ACCEPTABLE QU'UNE ARME AUTONOME PUISSE TUER UN ÊTRE HUMAIN ? UN ANIMAL ?

Une très forte majorité des gens répondent non à cette question, souvent avec beaucoup d'emphase et de points d'exclamation. On invoque que « tuer doit rester dans les mains de l'humain qui doit avoir pleine conscience de son geste ». L'idée d'interdire légalement les systèmes d'armes autonomes est d'ailleurs mentionnée plusieurs fois. Un répondant souligne aussi le risque d'une course à l'armement et la possibilité d'erreurs de programmation. Quelques répondants font une distinction : non pour un humain, oui pour un animal (« dans un contexte de régulation de population »). Certains cas semblent

faire exception : une machine tuant un condamné à mort ou un « tigre qui s'échappe de sa cage et met en danger le public ». Dans tous les cas, il apparait que l'IA ne devrait être qu'un outil de mise à mort dont la responsabilité incombe à l'humain. Un répondant développe toutefois une réflexion critique et soulève une question pertinente : « si cette arme peut prendre une meilleure décision qu'un humain, pourquoi pas ? ».

On remarque aussi que cette question dépend beaucoup du contexte : "It would be acceptable for an autonomous weapon to kill a human being or an animal in any circumstances where it would be acceptable for a human or other creature to kill a human or animal."

EXTRAITS CHOISIS

« Les armes autonomes ne devraient pas exister, elles devraient être bannies au même titre que les armes chimiques. Un humain devrait toujours être aux contrôles d'une arme : il aura ainsi la responsabilité morale de son geste. »

"No! (...) a horrific scenario could ensue from an unethical manufacturer or rogue programmer who perhaps, unbeknownst to either the weapon's company or weapon purchaser, may secretly design, code & program autonomous weapons which reflect their secret views & biases as a Neo-Nazi or KKK supporters, for example."

"Why would you think a HUMAN should be able to kill somebody? If you have a reason, then why doesn't it apply to an AI. There's no reason humans should always occupy a privileged position with respect to killing other humans. Obviously "AI" at the moment is not even ready for consideration for this, but that's unlikely to be permanent (assuming you think anything or anybody should be killing whoever or whatever you're thinking about killing). The interesting question may be how you'll know when it's changed, and how you manage the transition."

3. EST-IL ACCEPTABLE QU'UNE IA CONTRÔLE UN ABATTOIR ?

Comme à la question précédente, une large majorité répondent que non (on note toutefois plus de réponses favorables). On retrouve l'argument de la banalisation de la violence par la distance psychologique invoquée dans la question précédente : « Cela éloignerait encore plus l'homme de cette action de tuer l'animal », ou encore : « Il ne faut pas offrir une nouvelle couche de couardise à l'humain derrière laquelle il peut se cacher en déléguant une tâche moralement répréhensible à un robot. » L'argument environnemental est aussi invoqué : « Ce n'est pas l'orientation à privilégier pour l'avenir de la planète et des humains qui l'habitent. Je préfèrerais que l'IA contrôle des serres et des édifices zéro émission de CO2 et de déchet. »

Quelques arguments en faveur d'un tel projet : éviter la maltraitance, le stress des animaux et améliorer l'hygiène. (Un répondant explique néanmoins qu'un humain devrait toujours superviser les opérations d'abattage précisément pour éviter la cruauté...). Une position mixte : l'IA pourrait contrôler le découpage et l'emballage, mais pas l'abattage. Plusieurs se demandent toutefois si les abattoirs sont acceptables, même sans IA.

EXTRAITS CHOISIS

« Je comprends que cela soulagerait les responsables de ces tâches morbides cependant cela ne serait absolument pas éthique. »

« Les conditions d'abattage seraient peut-être légèrement améliorées, mais la pratique elle-même durerait plus longtemps, puisque l'on pourrait plus facilement s'en détourner. »

"Yes, to the extent that the AI follows humane (and human) protocol."

"Interesting question. One side of me says yes, the other no. What we are doing to other animals that we raise for food already has some serious ethical issues. When I read about the life of the average chicken raised for food, I was shocked. Totally automating the process of raising food, including having AI do the killing would just put the fate of these animals even more out of sight, out of mind. So, on balance, I think I am against an AI controlled abattoir."

"Slaughterhouses already exist and won't stop existing anytime soon. AI can make sure that the method of slaughter is ethical and is done in the most humane way possible. This can also strictly ensure and maintain safety standards."

4. DEVRAIT-ON CONFIER LA GESTION D'UN LAC, D'UNE FORÊT OU DE L'ATMOSPHÈRE TERRESTRE À UNE IA ?

Cette question suscite beaucoup de méfiance à l'égard de l'IA, mais aussi des espoirs de solutions à la crise environnementale. Il ressort de cela encore une fois que l'IA qui prendrait soin de l'environnement devrait être « configurée par des êtres humains responsables qui ont à cœur la préservation ». Certains y voient même un espoir, en particulier pour le climat, mais dans une logique de coopération humain/IA plutôt que d'une délégation du problème à l'IA. Plusieurs répondants espèrent une IA qui ne serait pas corruptible et dans une logique de recherche de profit.

On note le risque de détournement malveillant d'une IA qui aurait une telle mission. Et une pointe de cynisme : « l'humain détruit tous les milieux naturels auxquels il touche ou presque, alors ça ne pourrait pas être pire… » Surgit aussi le thème du remplacement de l'humain : « On pourrait tout laisser faire à l'IA, mais reste à savoir si l'on souhaite que l'homme devienne assisté sur tout son potentiel. » Et un principe démocratique : aucune entité humaine ou artificielle ne devrait pouvoir décider seule de la gestion de l'environnement — cela devrait venir d'une coopération de tous les humains.

EXTRAITS CHOISIS

« Non, car nos connaissances sont largement insuffisantes pour juger des répercussions à long terme des actions décidées par l'IA. »

"It depends on what the instructions given to the IA and how much absolute control it holds. I think an AI trained on environmental systems and with ability to monitor and consider big (environmental) data could make much better decision than any group of individuals, effectively helping to protect the environment and regenerate those

that may have been affected by industry, etc."

« Cela peut se faire avec l'assistance d'une IA ; mais l'IA ne sait pas d'elle-même ce qui est bon pour le lac, la forêt ou l'atmosphère : elle est plus performante du point de vue de la rationalité instrumentale, mais ne peut se donner à elle-même ses propres fins. »

"At present, AI would be most useful in the collection and analysis of data."

"Eventually, machines may be more competent than people to make almost all decisions. But, if we give the machine control and stop monitoring what and how it does what it does, the ability of human beings to manage our affairs will pass out of the living memory of humans, and we will be entirely dependent upon machines. This does not seem to me to be a good future for human beings."

DEVRAIT-ON DÉVELOPPER DES IA CAPABLES DE RESSENTIR DU BIEN-ÊTRE ?

Beaucoup d'hésitations et d'intuitions contradictoires ici. Est-ce seulement possible, techniquement parlant ? On mentionne que la sentience pourrait permettre à l'humain de contrôler ou de punir des IA. D'autres y voient un intérêt afin qu'une IA puisse mieux comprendre les humains (et autres êtres sentients) et faire de l'empathie avec eux. L'intelligence émotionnelle semble aussi requise pour faire de bons jugements moraux. Toutefois, une empathie simulée parait suffisante : car on voit un danger à ce qu'une IA sentiente privilégie son bien-être personnel aux fonctions que l'humain lui aurait assignées. « L'IA doit rester un instrument au service de l'humain, pas un simili-humain ». Et plusieurs se demandent « à quoi bon ? » Et un répondant s'inquiète pour l'IA : « Je préfère que l'IA comprenne le bien-être plutôt qu'elle le ressente, surtout que dans la notion de bien-être il y a aussi la notion de mal-être. »

EXTRAITS CHOISIS

"I think it makes most sense to approach the development of general AI as the development of a calculator/tool. Developing a personified, sentient AI may be bringing new life to the world. I'm sure it would be treated fairly or with rights."

« Question des plus complexes. Si elle ressent du bien-être, cela va générer le désir de le maximiser. Utile dans une logique de récompense des apprentissages. Toutefois, va-t-elle balancer son bien-être de machine avec celui des humains et des êtres vivants ? »

« Oui, mais seulement en fonction de l'accomplissement des tâches qui lui sont dévolues. Il serait ainsi possible de développer des IA qui, grâce à cette satisfaction du devoir accompli, chercheraient à s'améliorer constamment, mais seulement dans le domaine spécifique où elles opèrent. »

AUTONOMIE

1. COMMENT L'IA PEUT-ELLE CONTRIBUER À L'AUTONOMIE DES ÊTRES HUMAINS ?

l'IA présente une relation ambiguë avec l'autonomie : elle nous rend dépendants d'elle-même (« on ne pourra plus se dissocier de l'IA), tout en libérant l'humain de certaines tâches cognitives aliénantes (ex. : conduire une voiture, démarches administratives), voire de la nécessité de travailler. Dans les commentaires, c'est toutefois l'aspect positif et libérateur qui ressort le plus. Le modèle du partenariat est une option, tout comme celui de l'IA simple assistante. Et de grands espoirs sont possibles : l'IA pourrait améliorer la condition humaine et tout particulièrement les personnes en situation de handicap. On note aussi l'espoir d'une médecine moins invasive et qui vienne en aide aux personnes âgées en perte d'autonomie.

EXTRAITS CHOISIS

« L'IA devrait être utilisée pour redonner de l'autonomie (physique ou mentale) à des personnes handicapées. Seule une personne apte à contrôler entièrement la configuration des algorithmes d'un système d'IA pourrait gagner de l'autonomie, tous les autres en perdent parce qu'ils se fient à des décisions qui sont prises par quelqu'un d'autre. »

« Aucun système n'aidera (et n'aide aujourd'hui) l'autonomie des humains s'il vient de l'entreprise privée. La règlementation et l'implication du secteur public sont essentielles au maintien de l'équilibre. »

« En les libérant des tâches qu'ils ne souhaitent pas faire, et en améliorant leur état émotionnel et leur compréhension du monde. »

"HUMANs will deliberately develop AI to force other humans to follow their values or act according to their interests. And those humans will see themselves as benevolent in doing that which is the really scary part."

"AI can automate most of the trivial things that we spend a lot of time doing. Almost everything that we do without actively thinking about it can in a way be simplified or made more convenient using AI. But this also has to ensure that humans don't become too dependent on the technology, which would then handicap their life instead of providing more autonomy."

2. FAUT-IL LUTTER CONTRE LE PHÉNOMÈNE DE CAPTURE DE L'ATTENTION DONT S'ACCOMPAGNENT LES AVANCÉES DE L'IA ?

Le phénomène suscite beaucoup de scepticisme (« j'ai besoin de plus d'information »). Mais plusieurs notent le risque d'une « hypnose technologique », « surtout chez les adolescents ». Un répondant résume : « il ne faut pas devenir les esclaves de nos technologies ». Et un autre propose de soigner l'IA avec de l'IA : « Il faudrait savoir dans quels buts est captée l'attention. Les buts mercantiles, qui sont le moteur d'applications comme Facebook devraient pouvoir être freinés par d'autres applications IA, sortes de contre-mesures, mises à la disposition des utilisateurs pour lutter contre les intrusions. »

Évidemment, capter l'attention des gens sur des problèmes éthiques parait une bonne idée : « c'est

un moyen comme un autre de s'assurer que ces débats auront lieu. » Et une remarque pleine de bon sens : "Yes, businesses should be prevented from manipulating people's attention in ways that those people don't control or understand. That's not intrinsically related to AI; it's just that AI is a convenient, powerful, and therefore dangerous tool for it."

EXTRAITS CHOISIS

« Oui. En éduquant les populations dans un premier temps. Puis en légiférant pour imposer un cadre opérationnel s'inscrivant dans des valeurs humanistes (vérité, justice, bonté, respect, etc.) »

"All technologies, from radio frequencies, nuclear energy to cryptography must live within a regulatory framework. Attention seeking AI could be classified as addictive entertainment, like gambling."

"One way to combat this is awareness about the problem, the fact that this is happening is not known to many (hypothesis). And give the user proper tools to combat this: nudge the user to actually learn the skill using small dopamine hits until the user doesn't need it anymore."

3. FAUT-IL S'INQUIÉTER DE CE QUE DES HUMAINS PRÉFÈRENT LA COMPAGNIE DES IA À CELLE D'AUTRES HUMAINS OU D'ANIMAUX ?

Aucune tendance ne se démarque de façon claire. Certes, on s'inquiète de ce que la technologie sépare ou isole les humains : « l'humain doit rester un être social » et il faut prendre garde à ce que l'humain n'oublie pas ses compétences sociales comme l'empathie. Mais de ce point de vue, l'IA « ne serait pas pire que les jeux électroniques ». Des études de psychologie seront sans doute nécessaires pour évaluer les risques d'une nouvelle dépendance. Mais on y voit aussi un bénéfice potentiel pour les personnes seules ou pour certains profils psychologiques : des enfants autistes, par exemple, peuvent avoir plus de facilité à communiquer avec une IA qu'avec un humain. La chose devrait toutefois rester marginale, car « si un humain ne veut plus de contact avec d'autres humains, l'humanité disparait ». Mais attention au paternalisme : si cela ne nuit pas aux autres, pourquoi empêcher des relations fortes entre humain et IA. Et on accepte en général que certaines personnes préfèrent la compagnie des animaux à celle des humains.

EXTRAITS CHOISIS

« Si les agents IA n'ont pas de sensibilité ni de sentiments, ils ne sont pas de bonne compagnie. Moins que des animaux même. Ce sont des choses, pour l'instant. Des machines. »

« Si la personne n'a pas d'autre option, ça peut être une bonne chose. Sinon on va commencer à avoir des difficultés à vivre en communauté. »

« Non, beaucoup d'êtres humains sont déjà captifs de relations avec des objets ou encore avec des personnages fictifs (télévision,

téléromans, "amis" des réseaux sociaux). La compagnie d'une IA aurait au moins l'avantage de présenter un certain degré d'interactivité qui pourrait s'avérer particulièrement bénéfique chez les personnes âgées ou seules. »

"This is a legitimate concern. It can be compared to the preference for texting as a substitute for direct human-to-human interaction."

"If you care about people's autonomy, then LET THEM MAKE THEIR OWN DECISIONS. It doesn't matter whether you're "worried", because it's purely none of your business, full stop."

"Even if technologies like VR [virtual reality] are developed to an almost realistic level, it would only increase social isolation, and it would be detrimental in the long run. Social security, i.e the fact that there are people to support you and will be there with you in your time of need, is invaluable!"

4. PEUT-ON DONNER SON CONSENTEMENT ÉCLAIRÉ FACE À DES TECHNOLOGIES AUTONOMES DE PLUS EN PLUS COMPLEXES ?

Cela va être difficile pour deux types de raison, soulignent plusieurs répondants : la complexité des machines et la complexité des clauses légales. Personne ne lit les termes de consentement des applications ou des plateformes qui sont trop complexes (jargon juridique) : lorsque les gens « acceptent » ont-ils réellement le choix ? Ces consentements ne peuvent pas être considérés véritablement éclairés. "How often do we sign off on online agreements saying we read them when we didn't?"

Il reste donc à créer des systèmes qui donnent confiance et soient sécuritaires. Le manque de littératie numérique est aussi pointé, ainsi que la nécessité d'y remédier par l'éducation. Cela montre d'ailleurs « l'importance d'établir un code d'éthique sur lequel l'IA se constituera ». Une solution pourrait venir de l'IA elle-même : elle devrait être capable de répondre à nos questions pour nous aider à former un jugement éclairé. Mais un autre danger guette : "Information presented to humans will naturally inform (and bias) decision-making. Humans are quick to assume that algorithms or information provided by statistical analysis is somehow void of bias."

EXTRAITS CHOISIS

« Il serait bien d'encadrer juridiquement cette notion vis-à-vis des entreprises et des organismes publics faisant des affaires au Québec. »

« C'est impossible. La seule chose à faire est d'établir ou de rétablir une confiance avec ceux qui construisent et sont propriétaires de ces technologies par un contrôle social et politique qui satisfasse le plus grand nombre des utilisateurs, en réduisant les abus et détournements que les créateurs et propriétaires de ces technologies pourraient être tentés de réaliser. »

« Probablement pas. Je crois qu'il est déjà impossible de donner un consentement éclairé pour des technologies informatiques qui ne se base même pas sur l'IA. Par

exemple, comment être sûr que nous ne sommes pas espionnés par les logiciels que nous achetons. »

"For decently complex systems, the user has to be fully made aware of how the data being generated can and might be used, along with theoretical guarantees or open code base proving their claims. But for very complex systems, here, even the creator wouldn't know how the data might be used completely. But, even in the worst of the cases, rigorous proof of claims and possible benefits, analysis on a test group can help earn the trust of the user and allow the person to give consent."

"As technology advances, the demands for our consent will increase exponentially. Under those conditions, the unaided human will not be able to give truly informed consent in many of the cases where it is demanded. The proof is that we already have become conditioned to signing off on agreements that we have not actually read or understood. The demands are only going to increase. The solution, if there is one, would involve "loyal" AI agents assisting us."

5. FAUT-IL LIMITER L'AUTONOMIE DES SYSTÈMES INFORMATIQUES INTELLIGENTS ? UN HUMAIN DEVRAIT-IL TOUJOURS AVOIR LA DÉCISION FINALE ?

Beaucoup de réponses positives. L'être humain doit toujours être aux commandes, garder la main. L'IA est un outil, une aide à la décision. Les points de vue divergents sont toutefois intéressants : « L'IA est potentiellement plus précise, moins biaisée et bientôt plus créative que l'humain. Profitons-en ! » et dans le même ordre d'idée : « Les humains sont corruptibles. Une IA peut avoir un code moral plus strict que les humains. » Il pourrait aussi y avoir des contraintes liées au fait qu'on a besoin d'une prise de décision urgente.

Le contexte est évidemment important : faire des croissants ou lancer une attaque, ce n'est pas la même chose. L'humain devrait minimalement pouvoir prendre la décision d'arrêter un système autonome. Et cela ne semble pas négociable « dans le cas de décisions complexes incluant une dimension éthique engageant la responsabilité ».

EXTRAITS CHOISIS

« Il pourrait arriver un point dans le développement des systèmes où il sera possible de démontrer qu'un être humain n'a pas la capacité de prendre une meilleure décision que l'ordinateur. »

« La décision finale, non, car l'avantage de l'IA est de pouvoir prendre une décision instantanée en fonction d'une somme de paramètres qu'un humain ne pourrait jamais analyser aussi rapidement. Mais la responsabilité de la décision, elle, doit toujours être assumée par un humain. »

« Oui et oui, les systèmes informatiques sont des aides à la décision et doivent le rester. Pourquoi donner à un cyborg le pouvoir sur nous ? »

« Les décisions fondamentales doivent être humaines et reposer sur le plus large consensus possible. »

"You always want to have the option of an off switch. And we need to build systems in such a way that we can come to an understanding of how the machine is making the decision."

"Obviously with the current state of the technology, you can't let it have total control over too many things. That is unlikely to be true forever; eventually the AI is probably going to be smarter than the human... and possibly more benevolent than the human, which is where you should really be putting your energy. At some point the question may be whether the human should even get any input into certain decisions, especially into decisions that affected more than just that human."

"If the human does NOT always make the final decision, then there needs to be a transparent interface so that users can correct the decision-making computer system when it makes mistakes (like google translate, you can provide a better translation)."

JUSTICE (ÉQUITÉ, SOLIDARITÉ, DIVERSITÉ)

1. COMMENT S'ASSURER QUE LES BÉNÉFICES DE L'IA SOIENT ACCESSIBLES À TOUS ?

Par un prix abordable (ou la gratuité), par l'*open source* et en exposant clairement quelles sont les décisions que l'IA prendra à notre place (transparence). Mais est-ce possible dans le système capitaliste que nous connaissons ? « Le secteur privé ne devrait pas pouvoir exploiter une rente à son seul profit et au détriment du reste de l'humanité. » On pourrait d'ailleurs taxer les compagnies qui s'enrichissent excessivement grâce à l'IA (cela nuirait-il à l'innovation ?).

L'éducation pourrait avoir son rôle à jouer pour lutter contre la fracture digitale. C'est le rôle des gouvernements (voire de l'ONU) que de répartir équitablement ces bénéfices et de s'assurer que les valeurs de l'IA soient alignées avec les valeurs humaines. Un *basic income*, un appel au réalisme politique tempère les attentes : « ne soyons pas utopistes, ce n'est pas l'IA qui crée les inégalités, c'est l'humain ». Un répondant note aussi que les technologies de l'information rendent possible une démocratie participative. Un autre évoque le *basic income*.

EXTRAITS CHOISIS

« Construire les IA dans l'intérêt commun plutôt que comme propriété privée. Réglementer pour forcer les formes avancées d'adopter une licence libre GNU par exemple et de privilégier le partage de l'information. »

« Il ne faut pas laisser l'IA entièrement à la merci de l'entreprise privée. »

« Les avancées possibles des AI, par exemple la découverte d'une nouvelle protéine, doivent être des biens collectifs. »

"General quality of life for everyone should be improved with AI. Legal system seems to be one that will be greatly affected and see a lot of change, for the better."

« La fracture de l'accessibilité pourrait dépendre de la mainmise ou non de grands groupes privés sur les données générées par la population. »

« Il faut revoir en profondeur les lois internationales sur les brevets. Le développement des IA ne progressera véritablement que si l'information qui les sous-tend est du domaine public. L'appropriation de cette technologie par des groupes d'intérêts spécialisés (corporations, armée, gouvernements) ne doit pas être rendue possible, sans quoi elle sera inévitablement détournée pour servir ces intérêts plutôt que les citoyens. »

« Faire de l'équité un chantier central. Inclure les chercheurs et les groupes communautaires qui exercent la collaboration dans le design de solutions équitables. Voir les travaux de l'Unité soutien (SRAP) et du chantier Mobilisation et participation citoyennes d'Alliance santé Québec. »

"Give free Wi-Fi to the poor for starters."

"This is a very complex question. One could argue that everyone already benefits from AI through "free" products like Facebook and Google Maps. What is missing is an understanding of the market value of someone's data relative to the machine's ability to build a more powerful model. Governments at all levels need to be using AI with the data they currently manage as another part of their policy-making tool set."

2. FAUT-IL LUTTER CONTRE LA CONCENTRATION DU POUVOIR ET DE LA RICHESSE AU SEIN D'UN PETIT NOMBRE D'ENTREPRISES EN IA ?

Les réponses sont nettement positives. En favorisant l'*open source* et les licences libres GNU. Car c'est l'État plutôt que le secteur privé (les GAFA) qui a la confiance des citoyens. Les inquiétudes sont réelles : « La démocratie survivrait-elle avec une IA prédominante dans de mauvaises mains ? » Comment faire, toutefois ? On n'est même pas arrivé à faire que le logiciel libre supplante le logiciel propriétaire. Nationaliser pour rester « maître chez soi » ? Quoi qu'il en soit, l'IA devrait être vue comme un bien commun qui ne sert pas une minorité. Un répondant évoque la nécessité d'un organisme antitrust pour briser certains monopoles.

Cependant certains valorisent un modèle plus concurrentiel : « Si certaines entreprises parviennent à se créer une niche qui leur rapporte pouvoir et richesses, grand bien leur fasse. Mais la connaissance doit être du domaine public afin de favoriser la concurrence. » Un répondant propose de faire des données personnelles la propriété des individus qui pourraient se prévaloir également d'une IA d'assistance personnelle loyale envers eux.

EXTRAITS CHOISIS

« Évidemment, il faut lutter contre la concentration du pouvoir, point. »

"Yes, it seems there will be a lot of power available to those who control AI systems. New legislation/law will be required to monitor this, along with taxes on automation, etc."

« Il faudrait surtout que les programmes de base soient universels et bâtis pour le bien commun. Sinon ce ne seront que des robots au profit de ceux qui dirigent déjà malicieusement le monde dans leur propre intérêt, alors soit ça ne changera rien, soit ça empirera les inégalités, la violence, les conflits, etc. »

"The hands of a small number of AI companies or the hands of a small number human entities (i.e. the 1%) should not have more power and wealth than the 99% of human beings on earth. Powerful entities should adopt socially responsible behaviours at all time, especially when in presence of the public. (...) The democratization of AI should definitely empower the 99% of human beings."

3. QUELLES SONT LES DISCRIMINATIONS QUE L'IA POURRAIT CRÉER OU EXACERBER ?

Toutes les formes de discrimination « classiques » semblent pouvoir être exacerbées par l'IA, en particulier les discriminations « sociales, raciales, économiques », mais aussi « linguistiques et culturelles ». Entre les gens, mais aussi entre les groupes ou entre les États. Un scénario dystopique se profile : celui où une nouvelle classe d'ultra-riches (le 1% ?) utilise l'IA pour perpétuer les inégalités socio-économiques. On mentionne aussi que l'accès à la technologie peut être exclusif et excluant, en particulier pour les personnes plus âgées.

Un répondant précise le type de mécanisme que pourrait encourager l'IA : « L'IA peut être le bouc émissaire parfait sous forme d'une BLACK BOX : Pourquoi je n'ai pas reçu une marge de crédit M. le directeur de la banque ? Ah, c'est le système qui nous a donné le résultat, je suis désolé. » Il semble aussi clair pour les répondants que ce sont les humains en tant qu'individus ou en tant que groupes (ex. racisme systémique) qui sont et seront responsables de ces discriminations — pas l'IA.

EXTRAITS CHOISIS

« [Il faut se méfier de l'] apparition d'une "caste" des experts en IA, connus ou occultes, détenant le savoir, donc le pouvoir. [Il faut aussi se méfier des] discriminations selon l'état de santé (flirt avec l'eugénisme), discriminations raciales, sexuelles, envers les gens âgés, envers les femmes, etc. [Attention enfin aux] discriminations économiques, accentuant la pauvreté du plus grand nombre et le pouvoir des riches sur les décideurs. »

« Il y en a trop... Là est justement le problème. Nous avons de la difficulté à établir ce qu'est

une discrimination ou si nous en faisons déjà. Comment l'IA pourrait le déterminer à notre place sans justement utiliser les mêmes discriminations que nous lui fournissons en lui donnant des datas. »

« Les réseaux sociaux sont déjà source de stéréotypes, propos racistes, sexistes, stigmatisants. On peut penser à filtrer cela, ce que déplace le problème. Ces filtres aussi pourraient comporter des discriminations indues. »

« Les algorithmes devraient être développés par des équipes multidisciplinaires et multiculturelles afin de ne pas perpétuer de préjugés de genre, d'économie, d'ethnie, etc. »

« Si l'IA participe au bien-être et à l'autonomie, les personnes qui en ont besoin, mais n'y ont pas accès, seront d'autant moins bien loties. »

« Si l'IA est déployée par des groupes d'intérêts spécialisés (armées, gouvernements, corporations), elle ne servira que leurs intérêts du moment au détriment de la population. »

"See weapons of math destruction. AI models with labelled training data that is discriminatory will simply perpetuate and reinforce these discriminations."

*"It's going to be hard to deal with that, because in order to admit that the AI is going to find a regularity, you have to admit that the regularity exists. You have to be willing to say, "Yes, XXX people *are* more likely to default on loans, but we want to ignore that anyway". After that, it's a relatively simple technical problem to make the AI implement your wishes. Short-term AI, anyhow."*

4. LE DÉVELOPPEMENT DE L'IA DEVRAIT-IL ÊTRE NEUTRE OU CHERCHER À RÉDUIRE LES INÉGALITÉS ÉCONOMIQUES ET SOCIALES ?

La plupart des répondants sont favorables à une IA qui contribuerait activement à réduire les inégalités économiques et sociales. Plusieurs y voient même une priorité. Un répondant optimiste pense que les réductions des inégalités seront un effet mécanique du développement de l'IA. Un autre voudrait qu'elle promeuve surtout l'égalité des chances. Toutefois, quelques sceptiques préfèreraient qu'elle reste neutre : « Qui va indiquer quelles inégalités réduire ? » Et les plus pessimistes soutiennent qu'il y aura toujours des inégalités… ce qui n'empêche pas d'essayer de les réduire. Enfin, un répondant suggère que l'IA devrait rester neutre quant aux inégalités économiques et sociales lorsqu'il s'agit d'usages commerciaux, mais que les usages non commerciaux devraient viser davantage d'égalité.

EXTRAITS CHOISIS

« Oui l'intention devrait toujours être celle-ci, mais aussi la réduction des impacts environnementaux. »

« L'IA ne peut être neutre, donc autant assumer une direction meilleure pour tous. »

« Pourquoi développons-nous l'IA ? La réduction des inégalités ne me semble pas être la raison première ; cela ne veut pas dire cependant que le développement de l'IA devrait être neutre : les inégalités économiques et sociales pourraient faire office de "site contraint", pour que le développement ne se fasse pas au détriment de valeurs importantes. »

"AI models should be applied within a policy framework. No information system is neutral and any architect or policymakers must embrace the ethical challenges and opportunities when applying AI. In this context, reducing existing inequalities is a moral imperative. Machine learning models need to be conceived inside of a larger pipeline that can mitigate regressions and provides recourse for error."

"But we should make sure that by doing so we are not actively causing friction between different groups or trying to homogenize them. The effect, in that manner, should be neutral."

"It should be neutral in commercial settings, otherwise the technology might never be adopted at all—leading to no benefit to the society. But it should also reduce socioeconomic inequalities in a non-commercial setting by giving everyone access to the same tools and opportunities."

5. QUELS TYPES DE DÉCISIONS DE JUSTICE POURRAIT-ON DÉLÉGUER À UNE IA ?

Il ressort des contributions qu'aucune décision importante ne devrait être déléguée à une IA. L'IA ne doit être qu'un outil d'aide à la décision. Elle pourrait ainsi « accélérer le traitement des dossiers », voire « prendre des décisions faciles après une analyse des preuves », comme des décisions liées au paiement des contraventions.

L'IA pourrait être bénéfique dans d'autres aspects de la justice : « Détecter un mensonge ou un faux souvenir. Détecter les risques de récidives. » Si une intelligence artificielle générale était développée, alors on pourrait envisager que des IA remplacent des juges ; mais cette option est très loin de faire consensus, même s'il est avéré que les juges humains sont souvent biaisés dans leurs jugements et soumis à diverses pressions. Peut-être faudra-t-il un jour repenser l'institution judiciaire de fond en comble pour rendre possibles des « jugements artificiels ». Quoi qu'il en soit, la réduction des coûts et la démocratisation de la justice seraient une bonne nouvelle et l'IA pourrait certainement y contribuer, par exemple en facilitant l'accès à la jurisprudence.

EXTRAITS CHOISIS

« L'IA pourrait remplacer la personne qui prend des notes. »

« L'IA serait plus juste parce que non soumise aux émotions ou à la pression des médias et autres groupes d'opinion et de pression. La seule chose qui serait éventuellement à revoir serait le Code pénal, compte tenu des différences observées entre un

jugement humain et artificiel. »

"I don't believe any final decisions should be made by the AI. Seems the legal aid/technician and data processing could be best managed by AI."

« Les décisions impliquant le recours au jugement pratique complexe (jurisprudence) devraient être réservées aux humains. La justice est aussi un processus social. Ne l'oublions pas. »

« L'IA pourrait servir de recherchiste à la population (ainsi qu'aux juristes), en ayant accès à toute la jurisprudence. Ce travail démocratiserait l'accès à la justice puisque l'essentiel des coûts assumés par les citoyens est relié au temps passé par les juristes à faire ces recherches. »

"Current and near-future AI aren't going to be able to comprehend the law or apply it other than in cases so mechanical that you don't really need "AI" at all. I suspect that any real legal decisions will take a truly general intelligence."

"AI predictive technology can be used to help judges make better decisions. The idea is not to replace judges."

VIE PRIVÉE (INTIMITÉ)

1. COMMENT L'IA PEUT-ELLE GARANTIR LE RESPECT DE LA VIE PRIVÉE ?

Plusieurs répondants s'interrogent sur la pertinence de cette question : Comment l'IA peut-elle garantir ce respect ? L'impression est plutôt qu'elle la viole, à répétition et sans le consentement des utilisateurs. Il semble même y avoir une contradiction : l'IA a besoin de nos données pour se développer.

Mais il existe peut-être des options : « crypter tout », ne pas être invasif dans la demande de données personnelles. Quelqu'un remarque : « Le respect de la vie privée est garanti si la personne n'est pas exposée à une IA par défaut. » Il en va aussi de la responsabilité des utilisateurs : « C'est à chacun d'entre nous de contrôler son exposition : allez faire vos courses dans des boutiques indépendantes et payez en liquide, plutôt que d'acheter sur internet. »

On retrouve aussi une méfiance vis-à-vis du secteur privé : « Rien n'est garanti si c'est géré uniquement par le privé ». C'est pourquoi on appelle l'État et le législateur à la rescousse : il faudrait que les lois québécoises relatives à la vie privée soient respectées et améliorées. « C'est un gros défi », car ne serait-il pas déjà trop tard ? Nos données Facebook, par exemple, ont peut-être été siphonnées depuis longtemps par Cambridge Analytica ou une compagnie équivalente. Et c'est sans parler des « hackers ».

EXTRAITS CHOISIS

« Je crois que l'économie de l'information, basée sur la traçabilité, peut occasionner plus de partage d'information, mais en même temps plus de transparence dans leur usage, et donc avoir ses informations partagées n'aura pas de conséquences aussi lourdes si ceux qui la visionnent sont tracés aussi. »

"Let's face facts, there are, realistically speaking, no truly reliable guarantees that AI can respect people's privacy. Health records & private accounts are hacked all the time despite the best security upgrades that technology has to offer. Google reads our private e-mails, doesn't it?"

"Differential privacy—the idea that you can give away information about yourself without ever having it trace back to you as the source. But, if such a practice is possible and can be made prevalent then I believe that informed consent is possible. The user has to be fully made aware of how the data being generated can and might be used, along with theoretical guarantees or open code base proving their claims."

"Make people's private data truly their private property."

2. NOS DONNÉES PERSONNELLES NOUS APPARTIENNENT-ELLES ET DEVRAIT-ON AVOIR LE DROIT DE LES EFFACER ?

Les réponses sont massivement positives pour les deux sous-questions. Quelqu'un précise « et cela devrait être très facile comme procédure, pour que tous puissent le faire. » Un répondant s'oppose à l'idée que nos données nous appartiennent, mais cela n'empêche pas que nous devrions avoir « un droit de regard sur leur utilisation. » Si la majorité des répondants admet implicitement que ce sont les individus qui devraient posséder leurs données, certains l'envisagent plutôt comme un bien collectif.

L'effacement des données ne devrait toutefois pas entraver la justice (ou les services de santé) qui pourrait avoir besoin d'avoir accès à des données anciennes. Cet effacement ne devrait pas non plus causer de tort à autrui.

EXTRAITS CHOISIS

« Oui, chaque citoyen devrait être propriétaire de ses données privées au même titre que les artistes de leurs productions culturelles. »

« Non, mais les données devraient être considérées comme un bien national, comme les bibliothèques ou les réserves naturelles. »

« Absolument et de façon non équivoque. Seules les données essentielles pour le bon fonctionnement du gouvernement devraient être conservées : démographie, revenu, santé, judiciaire. Toutes les autres devraient pouvoir être contrôlées par l'utilisateur. »

"As long as companies own and licence IP, individuals should have a right to all data they create."

"Generally yes. But I have a very broad definition of what should be considered personal data (and should be private property). Within this larger view even our criminal records would be personal data that we own (though not without controls). It would be a category of personal data that we should not be able to delete—at least not whenever we choose."

3. DEVRAIT-ON SAVOIR À QUI NOS DONNÉES PERSONNELLES SONT TRANSMISES ET, PLUS GÉNÉRALEMENT, QUI LES UTILISE ?

Oui, répondent les gens à l'unanimité ! Quelqu'un précise : « Tout comme nous devons savoir qui entre dans notre maison, nous devrons savoir qui accède à nos données personnelles. » Un autre : « Oui, [et on devrait savoir] à qui, comment et dans quel but ». Un répondant remarque qu'on risque de se lasser de savoir qui utilise nos données et qu'on pourrait assez vite s'en désintéresser. Mais cela n'empêche évidemment pas qu'on ait le droit de le savoir.

EXTRAITS CHOISIS

« Oui, je pense que je devrais même avoir un portail où je contrôle 100 % de la donnée que je donne. »

« Nos données ne devraient jamais être transmises sans qu'au préalable une demande claire et concise ait été faite en ce sens. Pas de contrat de 20 pages en petites lettres où l'on doit deviner qu'il y a là une permission donnée *ad vitam aeternam*. Si l'on s'abonne à un service, l'information ne devrait jamais pouvoir être utilisée autrement que pour le service demandé. »

"Absolutely and they should be required to ask permission to do so on a regular basis. Permission is not granted in perpetuity."

"Absolutely. Personal data should be private property. We should defend it and allow the owner to control who can access it and to what extent they can access it. The current default—wherein we cede our data to others—is bad for citizens and bad for democracy. There is another option."

4. EST-IL CONTRAIRE AUX RÈGLES D'ÉTHIQUE OU D'ÉTIQUETTE QU'UNE IA RÉPONDE À VOTRE PLACE À VOS COURRIELS ?

Cette question soulève des intuitions contradictoires. Plusieurs remarquent que ce type de service existe déjà ou que certaines personnes ont des assistants humains qui répondent à leur place à leurs courriels. Une option serait que l'IA prépare la réponse, mais que celle-ci soit validée par l'humain (qui aurait ainsi le « dernier mot »). Un répondant précise que « l'important selon moi est que la personne qui utilise ce service ait la confiance et la compréhension nécessaires du service. » Une autre demande à ce que le procédé soit transparent, c'est-à-dire que l'interlocuteur sache que la réponse à son courriel provient d'une IA. Il n'y a peut-être pas de réponse générique à cette question : ça dépend des types de questions (« Es-tu disponible pour ce RV ? » vs « Penses-tu qu'on devrait embaucher telle personne ? »).

EXTRAITS CHOISIS

« Non du moment qu'il est stipulé de façon lisible que la réponse a été produite par une IA en lieu et place de l'utilisateur concerné. Si ce dernier veut utiliser une IA pour répondre à sa place, il en est de sa responsabilité… du moment que le fournisseur de service internet offre la possibilité d'activer ou de désactiver cette fonction. Il est bien entendu qu'il ne s'agit pas d'imposer un tel service. »

« Utile pour ceux qui ont à gérer un grand volume de messages similaires et peu complexes. »

« Ça dépend, si vous répondez toujours la même chose pour la même question, ça ne ferait pas de différence pour vous. »

« S'il est question de service à la clientèle, de répondre à un besoin humain à satisfaire qui engage la responsabilité d'autrui, je m'attends à ce que ce soit un humain qui réponde. »

"Yes. Human intent is a critical component to our society's framework. We can delegate to AI, but human dignity demands that you should know if you are interacting with a machine."

"Similarly, if an organization has a bot deal with people, it should always identify itself as a bot. People should always know if they are dealing with a human or a machine. And the organization that has bots dealing with people should always be held responsible for any actions the bot takes on the organization's behalf."

5. QU'EST-CE QU'UNE IA POURRAIT FAIRE EN VOTRE NOM ?

Une question ouverte qui suscite des réponses très diverses, allant de « rien » à « tout » (pour autant que l'on y a consenti). Entre les deux : programmer un rendez-vous, gérer mes finances, mon agenda, faire mes impôts et autres tâches administratives, voter (!). Mais je devrais toujours être tenu responsable des conséquences de ce que l'IA fait en mon nom. (Plusieurs répondants confondent cette question avec « Qu'est-ce que l'IA pourrait faire pour vous ? », par exemple : passer l'aspirateur).

EXTRAITS CHOISIS

« Tout ce que j'aurais préalablement approuvé. »

« Toutes tâches qui n'engagent pas une décision pour l'avenir. »

« Rien de sérieux pouvant avoir des implications légales ou émotionnelles. »

"Book appointments respond with numerical data that is already in the public domain, check on the well-being of family pets."

*"That depends on the AI. I wouldn't trust any *present* AI to do anything that I couldn't countermand or that people would interpret as a direct application of my personal judgment."*

"My recommendation is to adopt a paradigm in which each citizen owns private, 'loyal' AI tools (agent) that can help protect, manage, analyze and use a citizen's private data (stored in a protected online profile) to help that citizen at their

behest and only their behest. (…) Some people might say they can do simple repetitive tasks, perhaps review email. Others might allow their AI agent to browse the web to plan online shopping. Others might let the agent actually make purchases autonomously. Others might allow the AI agent to perform investment transactions for them. In an advanced future, some prefer to trust their AI to participate in a family vote about 'pulling the plug', given on its intimate access to its owner's private data, which could analyzing a variety information types taken from a personal profile, allowing it to use predictive analysis to help decide what the citizen might want if they were able to speak."

CONNAISSANCE (PUBLICITÉ, PRUDENCE)

1. LE DÉVELOPPEMENT DE L'IA FAIT-IL COURIR UN RISQUE À LA PENSÉE CRITIQUE ?

Les réponses sont contrastées, mais penchent plutôt pour le non. Du côté du « risque », on craint plusieurs choses : une perte de curiosité, la publicité, une normalisation de la pensée et la mise à l'écart des points de vue marginaux. Il se pourrait aussi que l'IA parle au nom des humains et qu'elle paraisse trop fiable : « La machine ne peut se tromper ; tout est dit ; il n'y a plus rien à ajouter. »

Du côté des avantages, plusieurs notent que le temps gagné par l'automatisation pourrait être investi dans la pensée critique et le fait que l'IA et les technologies de l'information rendent plus accessible l'information, voire qu'on pourrait programmer l'IA pour avoir une pensée critique — on perçoit aussi cette idée que l'IA pourrait être plus neutre que des humains. Enfin, on peut voir l'émergence de l'IA comme une belle occasion — ou une nécessité — pour les humains d'exercer leur pensée critique.

EXTRAITS CHOISIS

« Oui, mais pas si elle s'applique à faciliter la vie des gens en leur laissant plus de temps pour s'instruire et donc développer leur pensée critique. »

« Non, au contraire. La somme des connaissances humaines croît de façon exponentielle, au point qu'il devient impossible de connaitre tous les tenants et aboutissants d'un problème. L'IA, avec ses capacités de synthèse, permet aux humains de filtrer l'information redondante et de se concentrer sur l'essentiel. »

"I believe it certainly could compromise humans quest for knowledge & need to problem solve & therefore seriously impair our critical thinking & problem solving capacities & increase depression in people who may in future, have no motivation to use their god-given gifts & intelligence because they have been replaced by AI."

"It would definitely be more of a crutch than a tool if we become overly reliant on it. Instead the

development and the products that are created using AI tech should be such that it aids critical thinking, aids skill development and indirectly making life easier."

2. COMMENT MINIMISER LA CIRCULATION DE FAUSSES NOUVELLES OU D'INFORMATIONS MENSONGÈRES ?

Une question ouverte qui génère des pistes de solutions très diverses : soutenir financièrement les médias (locaux, traditionnels) qui vérifient l'information, investir dans le journalisme de qualité (qui multiplie les sources d'infos), éduquer les gens, utiliser une IA pour vérifier une info, punir ceux qui mettent de fausses nouvelles en circulation, les effacer, imposer des règles aux plateformes (type Facebook) qui font circuler ces fausses nouvelles. Notre accoutumance collective aux nouvelles « gratuites » (en un certain sens seulement) est aussi pointée du doigt.

Faut-il censurer les fausses nouvelles ? Un répondant prend position : « Il faut plutôt diffuser au maximum des articles de vérification des nouvelles, car la censure est contre-productive (elle peut par exemple alimenter des théories du complot) ». Un point de vue pessimiste : "It may become impossible as AI advances so too will its ability to mimic voices and fabricate images and video."

EXTRAITS CHOISIS

« Redéfinir le métier de journaliste. Développer un système d'accréditation des sources d'information. Reconnaître des experts en communication dans les différents secteurs de l'activité humaine. »

« Il y aura toujours de fausses nouvelles, il faut développer l'esprit critique et éduquer les jeunes en ce sens. »

« Il ne faut pas que la censure provienne directement de l'IA, par contre l'IA peut devenir un outil qui permet de prédire la probabilité pour qu'une nouvelle soit fausse. »

« Éduquer les gens à la pensée critique, à la recherche d'information crédible et à l'ouverture de leur conscience. »

3. LES RÉSULTATS DES RECHERCHES (POSITIFS OU NÉGATIFS) EN IA DOIVENT-ILS ÊTRE DISPONIBLES ET ACCESSIBLES ?

La réponse est sans ambiguïté positive. Et ce devrait être le cas pour tous les résultats de recherche dans tous les domaines, soutiennent plusieurs répondants. Ces résultats, précisent d'autres répondants, devraient être *open source* (notons qu'ils le sont déjà dans une très large mesure).

EXTRAITS CHOISIS

« Tout à fait. Et le plus possible, vulgariser ces résultats pour les rendre accessibles à tous. Pas de résultats opaques, avec des termes incompréhensibles... »

"This question has more to do with research than AI. Publicly funded research, with few exceptions, should be made available as a Social Good."

"Yes. I know people who think really powerful results should be kept from the "bad guys". That is a total pipe

dream. All you'll do by trying is to disadvantage the "good guys". Your best bet is to be open."

"YES!! Especially negative results. They would provide as much information, if not more about a particular problem."

4. EST-IL ACCEPTABLE DE NE PAS ÊTRE INFORMÉ QUE DES CONSEILS MÉDICAUX OU LÉGAUX SONT DONNÉS PAR UN « CHATBOT » ?

Pour les répondants au questionnaire, c'est largement le non qui l'emporte. Deux préoccupations semblent guider ces réponses ; le souci de transparence et celui de prudence : « Les conseils donnés par le chatbot peuvent être pris en considération de façon différente si la personne sait si elle parle avec un humain ou un chatbot. Un chatbot ne peut connaitre toutes les variables d'une situation. » Plusieurs remarquent qu'il est d'ailleurs facile d'informer une personne qu'elle communique avec un chatbot.

EXTRAITS CHOISIS

« Éventuellement oui. Aucun passager d'avion ne demande au maître de cabine si c'est le pilote ou l'autopilote qui contrôle l'avion. »

"The source of such advice being often critical to a person's well-being, one should be aware of the source of this information."

"No, every information should be presented along with the source exactly as it is, along with the analysis of how accurate or biased the information/advice might be. It may happen that the person may rely on that information even after realizing that it is from a chatbot, as it would getting good results. And that is the kind of relationship we'd like to foster."

5. EN QUEL SENS LES ALGORITHMES DEVRAIENT-ILS ÊTRE TRANSPARENTS QUANT À LEUR PROCESSUS DE DÉCISION ?

Cette question laisse beaucoup de répondants dubitatifs. La réponse qui revient le plus est « le plus possible » en ayant conscience des difficultés techniques en jeu ici (c'est-à-dire du problème de la « black box »). Si certains pensent que l'IA ne devrait tout simplement pas prendre de décision, les autres semblent d'accord pour qu'une IA prenne une décision, à condition d'avoir accès à une « justification déchiffrable par un humain ». Il se pourrait aussi que dans certains contextes, la transparence ne soit pas désirable. Plusieurs remarquent que la transparence sera importante pour susciter la confiance envers l'IA. Un répondant suggère de donner le degré de fiabilité d'une décision prise par une IA.

On note aussi que la transparence implique de savoir à partir de quelles données (ou type de donnée) une IA prend une décision et les valeurs (ou intérêts) qui guident sa décision.

Un participant suggère au contraire qu'on ne devrait pas être plus exigeant envers une IA qu'envers un humain.

EXTRAITS CHOISIS

« Une description du processus de décision des algorithmes devrait venir avec l'achat d'un produit IA, comme les modes d'emploi ou les garanties du fabricant qui accompagnent un produit lors de son achat. »

« Si les créateurs d'une IA ne sont pas capables de définir précisément la portée et les limites de capacité de décision d'une IA qu'ils proposent, celle-ci ne devrait pas pouvoir être commercialisée. »

« Les échelles de valeurs utilisées pour leur prise de décision. Voir les valeurs relatives des différents éléments décisionnels. Par exemple : Chat vs chien, collectif vs individu, etc. »

« Totalement transparent. Comment faire confiance si on ne sait pas sur quels principes ils se basent pour faire leur analyse ? Tout comme il est toujours pertinent de comprendre la méthodologie de recherche utilisée par des chercheurs. »

"You should be able to ask an AI why it made a choice then if you find its reasons lacking you should be able to make it change its behaviour."

"We may be able to infer decision-making processes but we should not assume that there is any internal motive or intent in an algorithm."

DÉMOCRATIE (PUBLICITÉ, DIVERSITÉ)

1. Faut-il contrôler institutionnellement la recherche et les applications de l'IA ?

La réponse est globalement positive, en particulier pour les applications de l'IA (la liberté de la recherche scientifique est une valeur importante). On suggère un « bureau de l'Ombudsman de l'IA » et des comités d'éthiques de l'IA ou encore une sorte de serment d'Hippocrate. On note aussi que « le sujet est trop éminemment politique et social pour être laissé aux mains du privé ». Ce contrôle, toutefois, ne devrait pas nuire à l'innovation (pour autant que celle-ci soit compatible avec le bien commun et les droits humains). Une difficulté inhérente à ce contrôle institutionnel relève de la politique internationale : comment des pays aux intérêts divergents pourront-ils se mettre d'accord sur des institutions communes ?

EXTRAITS CHOISIS

« Oui, à condition que nous développions une démocratie participative et que les gouvernements soient d'abord au service de la majorité, pas à celui du capital. »

« Non, mais poser des bornes est essentiel. »

"Yes but good luck getting China or Russia to follow along."

"Controlling AI research is simply not possible. The research itself should continue, but a broader communication framework explaining what AI can and cannot do is critical. Sensitizing researchers to the ethical ramifications of their work is also important (e.g. the Hippocratic oath).

2. DANS QUELS DOMAINES EST-CE LE PLUS PERTINENT ?

Question ouverte. Beaucoup répondent « dans tous les domaines ». La santé arrive largement en tête des domaines les plus souvent cités. On trouve ensuite (dans l'ordre) : l'armement, la justice, l'environnement, l'alimentation, la surveillance, la vie privée, la finance, la sécurité, l'éducation, le gouvernement. Sont aussi mentionnés : l'économie, l'industrie, l'épigénétique, le journalisme, le transport, les services municipaux, la recherche sur une super-IA (AGI), les voitures autonomes et les publicités ciblées.

EXTRAITS CHOISIS

« Dans tous les domaines liés à la vie (biologie) et à la vie en société. »

3. QUI DEVRAIT DÉCIDER — ET SELON QUELLES MODALITÉS — DES NORMES ET VALEURS MORALES DÉTERMINANT CE CONTRÔLE ?

Les répondants, souvent très indécis sur cette question, hésitent entre diverses options : le parlement, des consultations publiques, l'ensemble de la population (référendum, tirage au sort), un comité multidisciplinaire (experts, élus, citoyens), la commission d'éthique en sciences et technologie, un « comité de sages », une institution internationale (type ONU). L'idée que cette instance de décision devrait être indépendante (du pouvoir politique et économique) ressort à plusieurs reprises. On trouve aussi le souci que cette instance soit représentative de la diversité des citoyens.

EXTRAITS CHOISIS

« Je ne sais… Un comité conjoint, multidisciplinaire, universitaire, populaire et impartial. »

« Nous tous, en développant des modes d'information, de consultations et de prises de décisions qui impliquent le maximum de gens de toutes provenances. Pas la "démocratie" actuelle. »

« Bien des comités. Ceux-ci pourraient établir des règles, valeurs, etc., en lien avec chaque institution où il y aurait l'un de ces comités. Ainsi, ceux-ci pourraient établir une sorte de « charte », que devrait suivre l'institution et faire des recommandations… Qui bien sûr ne devraient pas être tablettées !

« Au Québec, la Commission d'éthique en science et technologie a déjà produit un document sur les villes intelligentes en indiquant les enjeux à considérer. D'autres projets IA pourraient être analysés par cette instance ou par d'autres instances gouvernementales spécialisées dans le domaine visé. Un ombudsman pourrait être nommé pour certifier des projets IA et recevoir des signalements liés au non-respect des principes de la Déclaration de Montréal en IA. »

« Comme l'IA touchera tous les domaines (loi, santé, science, société, arts), il est vital que des spécialistes de chacun de ces domaines soient représentés au sein de l'organisme. Le gouvernement doit financer

adéquatement l'organisme, mais ne peut intervenir dans son fonctionnement. De plus, un gouvernement ne devrait pas avoir le pouvoir d'abolir l'organisme ou d'entraver son travail. »

"Canadian's from all groups, backgrounds & beliefs."

"This should function like an IRB as in the drug development and testing industry."

4. QUI DEVRAIT CHOISIR LES « RÉGLAGES MORAUX » DES VOITURES AUTONOMES ?

Cette question suscite des réponses très variées : le parlement, une agence gouvernementale, l'État, les pouvoirs provinciaux, l'État en collaboration avec l'industrie, un comité d'expert en éthique, la SAAQ, le manufacturier qui porte la responsabilité de l'auto, une autorité de certification des logiciels, un comité d'usagers, les juges de la Cour suprême, un organisme international type ONU. L'usager pourrait aussi avoir le choix de certaines options. En passant, on constate aussi chez plusieurs répondants une défiance envers les voitures autonomes (« elles devraient être interdites »).

EXTRAITS CHOISIS

« Encore là, ça pourrait être des comités de citoyens. Il faudrait un représentant des piétons, un autre des personnes âgées, un autre des moins de 16 ans, un autre pour les vélos, etc. Chacun pourrait s'exprimer sur les réglages moraux à donner aux voitures autonomes. »

« Sûrement pas les compagnies qui les construisent ! »

« Un commissaire en éthique et le Bureau du Coroner au Québec. »

"It should be a multilateral decision (after thorough public discussion)."

"Judges/supreme court, whoever decides and uphold the existing ethical guidelines should have a major role to play in the decision. But along with them, community participation, transport businesses and authorities, AI researchers and developers."

5. FAUDRAIT-IL DÉVELOPPER UN OU DES « LABELS ÉTHIQUES » POUR LES IA, LES SITES WEB OU LES ENTREPRISES QUI RESPECTENT CERTAINS STANDARDS ?

Une forte majorité répond par l'affirmative, ce serait une « bonne idée », « un bon début ». Cela pourrait ressembler à une norme ISO. Un répondant se demande toutefois pourquoi toutes les entreprises et les sites web ne devraient pas respecter ces standards. Un autre précise : « oui au cas par cas avec une charte standard ». Cela soulève aussi un certain scepticisme : Ces certifications seront-elles respectées ? Ne risquent-elles pas d'être corrompues ?

EXTRAITS CHOISIS

« Des certifications sujettes éventuellement à révision de façon vigilante pour s'adapter à telle ou telle situation. »

"Communities are different, people are different. (...) We should make

sure that by doing so we are not actively causing friction between different groups or trying to homogenize them. The effect should be neutral."

"Definitely, at least three majors should be developed: corporate, government and individual ethical labels."

RESPONSABILITÉ (PRUDENCE)

1. QUI SONT LES ACTEURS RESPONSABLES DES CONSÉQUENCES DU DÉVELOPPEMENT DE L'IA ?

Les répondants identifient plusieurs acteurs responsables : les universités, les chercheurs, les entreprises, les éthiciens, les politiciens, ceux qui commercialisent les applications, le gouvernement, les décideurs économiques, ceux qui en tirent un bénéfice financier, les représentants élus, la société, les utilisateurs, chacun de nous. Mais c'est peut-être « les développeurs/créateurs, les entreprises et le gouvernement » qui reviennent le plus souvent. Certains font le parallèle avec les animaux domestiques ou les enfants : ce sont leurs propriétaires/tuteurs/parents qui sont responsables. Dans le cas des IA, ce pourraient être les propriétaires. On évoque aussi ceux qui testent les IA, qui autorisent leur déploiement.

EXTRAITS CHOISIS

« Les gens qui les fabriquent, les gens qui les distribuent, et si on peut les pincer, les gens qui les utilisent de façon malveillante pour nuire, blesser, tuer ou dominer les autres (incluant les animaux) ou dégrader l'environnement. »

« Tous les membres de la chaîne d'approvisionnement : du chercheur gradué, à la firme multinationale, en passant par les organisations de règlementation nationales, régionales et locales. »

« Les compagnies qui offriront les services devront être imputables et responsables, mais surtout les dirigeants de ces compagnies.»

"Whoever provides the results/ predictions of the AI decision-making. For example, Google is responsible for Google Translate."

"Researchers developing models are partially responsible, however the application of AI ultimately rests with the owner/operators."

2. COMMENT DÉFINIR UN DÉVELOPPEMENT PROGRESSISTE OU CONSERVATEUR DE L'IA ?

Une question qui reste souvent en suspens. Le développement progressiste rime avec collectif, transparence, moins d'écart de richesses. Le développement conservateur est assimilé à une certaine prudence : il ne faut pas se précipiter, il faut y aller graduellement. Quelqu'un remarque qu'il semble plus facile d'adapter la législation à l'IA que l'IA à la législation parce que le progrès va vite et qu'il semble difficile à freiner. Un autre que le développement progressiste devrait « favoriser les recherches alternatives ». Et un sentiment partagé par plusieurs : « On y va, on y va, peut-on faire autrement ? »

EXTRAITS CHOISIS

« En faisant des forums sur le sujet ! :-) Plus on en discutera,

de façon inclusive, plus il y aura un développement progressif de l'IA, dans le bon sens. Et aussi par l'éducation. Plus notre société sera éduquée, plus elle sera informée et fera des choix éclairés. »

« Pour le bien commun vs pour le bien privé. »

"It is progressive when it is maximizing freedom and agency. It is conservative when it is carefully monitored and cultivated as to i nsure safety."

"Conservative development: Checking, testing at each and every step. First in isolation, then within an isolated test group, and gradually deploy the AI."

3. COMMENT RÉAGIR DEVANT LES CONSÉQUENCES PRÉVISIBLES SUR LE MARCHÉ DU TRAVAIL?

Plusieurs idées reviennent : un filet social solide ou revenu universel (*basic income*), une réforme de la fiscalité avec une taxe sur les robots ou une meilleure répartition des richesses. C'est cependant le recours à l'éducation et de la formation qui est le plus souvent préconisé : les gens devront s'adapter, ce qui demandera plus de formation continue. La transition devra certainement se faire graduellement et être transparente : les gens devront être tenus informés. Mais tous ne sont pas inquiets : « Le marché du travail a toujours été en évolution et continuera ainsi ». Par ailleurs, plusieurs semblent espérer une libération du travail.

EXTRAITS CHOISIS

« Offrir un salaire garanti en échange de participation à l'information des communs [numériques]. »

« Le travail n'est pas l'horizon de l'humanité ni son but. Le temps libre dégagé et les gains de productivité générés devraient être mis en commun pour permettre à tous de moins travailler sans perdre en niveau de vie. »

« En redirigeant les personnes vers d'autres types d'emploi qui participeront à plus de cohésion sociale. »

« Besoin de ralentir la cadence ; il faudrait se donner des priorités qui visent d'abord à développer ce qui va être au service de l'humain avant ce qui va remplacer l'humain. »

"AI tax, job displacement compensation, basic living wage, and research/development of new jobs."

"The real cost of the introduction of AI technology is not just the money some people pay for it. It is the social, political, and economic costs—to everybody in society that need to be considered."

4. EST-IL ACCEPTABLE DE CONFIER UNE PERSONNE VULNÉRABLE AUX BONS SOINS D'UNE IA ? (PAR EXEMPLE, AVEC UN « ROBOT-NANNY »)

Les répondants sont très partagés sur cette question : « pour divertir, mais pas pour soigner », « pas sûr ». On semble craindre une disparition de l'humain dans le soin. On pointe l'importance de « la chaleur humaine » en particulier pour les personnes vulnérables. Reste que c'est mieux que rien : « Oui, s'il n'y a pas d'autres choix ». Il faut aussi voir que cela pourrait donner un accès à de meilleurs soins, en particulier lorsque les ressources humaines manquent. Plusieurs notent cependant le risque de se déresponsabiliser de nos devoirs envers ces personnes en les confiant à une IA. Le sujet est sensible et de tels robots devraient certainement être encadrés et supervisés.

EXTRAITS CHOISIS

« Pas totalement. Le robot-nanny devrait toujours être là comme complément au personnel humain. »

« Oui, si c'est possible de bien programmer l'IA pour qu'elle n'outrepasse pas certaines compétences plus sensibles. »

« À la personne vulnérable de décider. »

"No. The result could be disastrous as it has not been studied for decades to determine the social, psychological, mental & physical effects it would have on our children. It could also possibly make our children emotionally unable to connect & bond with their parents, siblings & other humans."

"Of course... consider how television is sometimes referred to as a babysitter."

5. UN AGENT ARTIFICIEL COMME TAY, LE CHATBOT « RACISTE » DE MICROSOFT, PEUT-IL ÊTRE MORALEMENT BLÂMABLE ET RESPONSABLE ?

Une question qui suscite plutôt des réponses négatives. On refuse de qualifier le *chatbot* comme raciste « puisqu'il ne comprend rien » pour faire porter la responsabilité sur ses concepteurs (Microsoft). Il n'empêche que « les conséquences de ses déclarations » pourraient avoir des effets bien réels. La plupart des répondants sont donc d'accord pour y voir quelque chose d'inacceptable. Un répondant explore l'angle juridique en envisageant de mettre les IA sous tutelle légale (comme des enfants ou des animaux) tandis qu'un autre les envisage comme de simples objets dont la responsabilité incombe au propriétaire.

EXTRAITS CHOISIS

« Non, je crois que nous devrions considérer les produits de l'intelligence artificielle comme si c'était des enfants. Il serait pertinent de leur donner un titre de personne n'ayant pas la personnalité juridique autonome complète. Comme cela, chaque produit intelligent artificiellement aurait un humain qui serait tuteur responsable de ses actes. »

« Il ne s'agit en fin de compte que d'un programme. Et l'on sait jusqu'à quel point des programmes peuvent être bogués, déficients et mal faits. »

« Pas pour le moment, la responsabilité vient avec la conscience, si l'IA n'est pas consciente, elle ne peut être blâmable. »

« Il est de la responsabilité du programmeur de s'assurer que son robot ne soit pas raciste et d'apporter tout changement requis dans les plus brefs délais. »

"We should accept that machine learning algorithms are non-deterministic and empower operators to explore their utility while being responsible operators."

"The responsibility (until proven that the being is actually sentient, if that's even possible) should be taken by: People who gave permission to deploy them > People who tested them > People who developed them. In that order."

"Humans are not good examples for AI agents. AI agents will be more efficiently learning from other AI agents than from human activities."

"No. I think it is always people who must be held responsible. I am against giving machines any kind of legal status similar to people. You cannot punish or hold responsible a machine. So, people must always be responsible."

3. SYNTHÈSE DES MÉMOIRES REÇUS

Une quinzaine de documents ont été reçus à la suite de l'appel lancé via le site web de la Déclaration de Montréal en novembre 2017 (avec une date limite de réception fin avril 2018). Il s'agissait de contribuer au contenu de la Déclaration, soit en discutant les 7 principes de la version préliminaire, soit en suggérant des recommandations concrètes. Ces documents vont du mémoire de synthèse de discussions collectives au texte d'opinion individuel. Ils sont écrits en français et en anglais ; on peut les lire sur le site de la Déclaration (cette synthèse ne rend évidemment pas compte de toute la richesse des mémoires reçus).

Les abréviations suivantes seront utilisées pour désigner les documents des organismes ou personnes suivantes :

AQT
pour l'Association québécoise des technologies

CAIQ
pour la Commission d'accès à l'information du Québec

MAIEM
pour le groupe Montréal AI Ethics Meetup

OIQ
pour l'Ordre des ingénieurs du Québec

SRAD
pour la soirée de réflexion autour de la Déclaration qui s'est tenue à l'UQAM

Hernandez
pour Annick, Guillaume et Raphaël Hernandez

McNally
pour John McNally

Musseau
pour Pierre Musseau-Milesi

Parent pour Lise Parent

Quintal et al.
pour Ariane Quintal, Matthew Sample et Eric Racine

Ravet
pour Jean-Claude Ravet

Robert
pour Bruno Robert

Wark
pour Grant Wark

VIE PRIVÉE

PRINCIPE PROPOSÉ :

« Le développement de l'IA devrait offrir des garanties sur le respect de la vie privée et permettre aux personnes qui l'utilisent d'accéder à leurs données personnelles ainsi qu'aux types d'informations que mobilise un algorithme ».

REMARQUES GÉNÉRALES

Le principe relatif à la vie privée est probablement celui qui a été le plus longuement commenté dans les mémoires reçus. La Commission de l'accès à l'information du Québec (CAIQ) en particulier, mais aussi le groupe Montréal AI Ethics Meetup (MAIEM), le rapport de la soirée de réflexion autour de la Déclaration qui s'est tenue à l'UQAM (SRAD), l'Ordre des ingénieurs du Québec (OIQ), Lise Parent (Parent), Annick, Guillaume et Raphaël Hernandez (Hernandez), Grant Wark (Wark), Quintal, Sample et Racine (Quintal et al.) ont proposé des recommandations explicitement liées à la vie privée.

Comme le remarque CAIQ, au Québec, la protection des renseignements personnels possède déjà des principes bien établis (RLRQ, A-2.1 ; la Loi sur l'accès, ainsi que RLRQ, P-39.1 ; la Loi sur le privé) que le développement de l'IA devra respecter : par exemple, les organismes qui collectent des données doivent déterminer par avance la finalité de cette collecte et en tenir informées les personnes concernées. On peut encore nommer les principes de nécessité, de consentement, de confidentialité, de destruction, de transparence, d'accès et de responsabilité (voir l'annexe de CAIQ).

Face à de nouvelles pratiques, on peut envisager au moins deux types de régulations : l'une coercitive, qui met l'accent sur les pénalités en cas de non-respect du cadre légal, et l'autre, préventive, qui vise à accompagner les changements de façon plus souple. Dans le contexte québécois, la CAIQ suggère le second type d'approche et insiste sur l'évaluation des risques en amont, sur l'utilisation de paramétrages qui soient, par défaut, les plus stricts possible, l'utilisation de technologie pour améliorer la confidentialité, la désignation dans chaque organisation d'un responsable à la protection des renseignements personnels qui soit imputable et la « transparence, au profit du citoyen ». On peut toutefois se demander si le rapport de force avec les grandes multinationales du numérique ne requerra pas aussi des mesures musclées de type plus coercitif que préventif.

Cette position peut faire écho à celle de l'OIQ (et de Parent) qui promeut la protection de la vie privée dès la conception (*privacy by design*) et propose de s'inspirer des bonnes pratiques existantes comme celles du règlement sur la protection des données (RGPD) récemment entré en vigueur en Europe.

Ce souci pour le respect de la vie privée s'accompagne souvent de celui pour la transparence. Le groupe MAIEM propose ainsi de bonifier le principe de vie privée en précisant que la transparence est essentielle — une analyse qui est aussi faite par CAIQ et SRAD. On retrouve donc ici la relation étroite entre les enjeux de protection de certaines informations (les renseignements personnels) et la possibilité de savoir qui détient quoi (l'accès à l'information), deux éléments qui mériteraient sans doute d'être plus explicites dans la Déclaration. On peut noter en passant que ces éléments peuvent entrer en tension : lorsque la transparence s'applique à des renseignements personnels qu'on voudrait garder confidentiels. Des arbitrages entre ces deux notions peuvent être nécessaires.

Il ne va d'ailleurs pas de soi que ces arbitrages soient consensuels, car comme le souligne MAIEM, les préférences en matière de vie privée peuvent « varier considérablement selon les cultures, les générations et les individus ». On constate en tout cas un consensus sur l'idée qu'il faut « préserver le contrôle du citoyen sur ses renseignements personnels et la gestion de son consentement » (CAIQ, SRAD). Quintal et al. s'inquiètent d'ailleurs de ce que la formulation initiale du principe de vie privée suggère que les données soient partagées par

défaut (le principe insiste sur la possibilité de savoir ce que deviennent les données personnelles, sans contester que ces données soient recueillies dans un premier temps). « La Déclaration, écrivent-ils, devrait inclure des garanties élevées pour la confidentialité des données des utilisateurs » [The Declaration should include improved safeguards for privacy of user data.]

SRAD note enfin que les techniques d'anonymisation des données ne sont pas encore matures pour être utilisables sans risque. SRAD remarque également le lien entre les enjeux de protection des données et les risques de discriminations algorithmiques. Mais cela ne signifie pas que des données protégées (par exemple le genre ou la race) ne devraient pas être recueillies dans la mesure où combattre une discrimination suppose habituellement d'avoir accès à ces informations.

PROPOSITIONS DE RECOMMANDATIONS

Le principe de vie privée, intégrant le souci de transparence, donne lieu à des recommandations plus spécifiques :

> Les gens doivent être informés, autorisés et capables de vérifier leurs renseignements personnels et la manière dont ils sont utilités à tout moment. [People need to be informed about, and allowed and able to check, their personal data and its uses at any time] (MAIEM).

> Il faut introduire une culture du « data privacy by default » comme c'est le cas en neuro-éthique, c'est-à-dire que, par défaut, les données personnelles ne devraient pas être partagées (Quintal et al.).

> Le « fardeau du consentement », c'est-à-dire le souci de s'assurer qu'il existe un véritable consentement libre et éclairé, devrait incomber aux entreprises/organisations et non aux citoyens, de même que celui de corriger des informations inexactes (CAIQ).

> Les gens doivent pouvoir comprendre comment leurs renseignements personnels seront utilisés (MAIEM, CAIQ, Hernandez, Parent).

> Les gens doivent pouvoir retirer leur consentement à l'utilisation de leurs renseignements personnels (MAIEM).

> Il faut rendre publics et ouverts les codes informatiques pour l'interprétation des résultats et les méthodes d'entraînement des algorithmes. (OIQ)

> Il faut sensibiliser les gens aux enjeux de protection de la vie privée (CAIQ).

> Les gens devraient pouvoir connaitre en tout temps la valeur monétaire de leurs renseignements personnels (Hernandez).

Enfin, une proposition originale et détaillée de Wark vient d'une certaine façon répondre à une interrogation de Hernandez : Comment créer un espace numérique privé ? Il s'agit en quelque sorte d'utiliser l'IA pour se protéger de l'IA.

> En effet, Wark suggère d'utiliser la technologie des « *smart contracts* » pour protéger les renseignements personnels et faciliter les échanges commerciaux et les interactions sociales. Il s'agit de développer un profil personnel sécurisé et une « IA loyale » qui feraient office de gestionnaire des données personnelles, répondant ainsi à nombre de défis identifiés précédemment. "For example, a loyal AI-agent must not adulterate its loyalty to its owner through overt or covert association with a business, such as an online store." Pour en savoir davantage, nous renvoyons au mémoire de Wark qui présente de façon détaillée à quoi une telle IA loyale pourrait ressembler.

Plusieurs mémoires évoquent enfin la mise en œuvre de ces recommandations. Du point de vue des politiques publiques, on peut envisager au moins trois manières de traduire ce souci pour le respect de la vie privée et de la transparence : par la réglementation, par l'autoréglementation ou par des incitatifs.

Aussi bien CAIQ que MAIEM sont d'avis que l'autoréglementation ne peut être suffisante. Il importe plutôt de moderniser la réglementation existante. L'un et l'autre (ainsi que Parent) insistent aussi sur l'importance de faire des audits des

entreprises et organisations. Cette modernisation peut prendre différentes directions : l'OIQ soutient « des mécanismes règlementaires souples », ce qui résonne avec l'approche préventive défendue par CAIQ.

On peut enfin envisager des incitatifs économiques pour favoriser les entreprises qui développent des technologies protectrices de la vie privée ainsi qu'une valorisation de celles qui font des efforts, par exemple au moyen de labels ou de certifications — une idée qui semble aussi avoir les faveurs de l'Association québécoise des technologies (AQT).

JUSTICE

PRINCIPE PROPOSÉ :

« Le développement de l'IA devrait promouvoir la justice et viser à éliminer les discriminations, notamment celles liées au genre, à l'âge, aux capacités mentales et physiques, à l'orientation sexuelle, aux origines ethniques et sociales et aux croyances religieuses »

REMARQUES GÉNÉRALES

Comme le principe de vie privée, celui de justice a aussi été discuté ou évoqué dans de nombreux mémoires : MAIEM, SRAD, OIQ, Hernandez, Parent, McNally, Ravet.

SRAD propose de distinguer les différents sens de la justice (avec Aristote) : la justice commutative qui règle les échanges entre personnes considérées comme égales et la justice distributive qui s'attache au mérite. Qui dans la société mérite quoi ? C'est surtout cette seconde acception qui semble mobilisée dans les mémoires reçus. Et elle soulève de nombreuses questions.

Peut-on identifier un principe de justice universel pour régir le développement de l'IA ? Ne faudrait-il pas se contenter de principes valables uniquement pour une communauté donnée ? Cette question délicate est au cœur de nombreux débats en philosophie politique.

MAIEM penche pour une approche non universaliste, ou du moins, qui essaye de tenir compte des variations culturelles et historiques de la notion de la justice : « Le développement et l'utilisation de solutions fondées sur l'intelligence artificielle devraient promouvoir la justice et l'agentivité humaine, définies de manière transparente par l'organisation qui définit le bien-être de la communauté cible (par exemple le gouvernement démocratiquement élu), de concert avec la communauté cible. Il devrait chercher à éliminer les inégalités et la discrimination au sein de cette communauté. »

[The development and utilization of AI-enabled solutions should promote justice and human agency as transparently defined by the target community's welfare-defining organization (e.g. democratically elected government), in concert with the target community. It should seek to eliminate inequality and discrimination within that community.]

On peut opposer à cette reformulation l'approche plus universaliste de Ravet qui trouve un principe universel dans l'idée (kantienne) de dignité humaine et dans celle de vie : « Les innovations en IA doivent se fonder sur le principe de non-instrumentalisation de l'humain et veiller à ne pas écraser la vie. » C'est aussi l'approche de SRAD qui, en plus de la notion d'égale dignité humaine, introduit l'idée de justice sociale : « Le développement de l'IA devrait promouvoir la justice sociale et respecter l'égale dignité humaine, notamment en visant à éliminer toute forme de discrimination, incluant celle liée au genre, à l'âge, aux origines ethniques, aux statuts sociaux, etc.».

Une manière d'articuler justice sociale et justice comme non-discrimination pourrait être de voir la première comme corrigeant des inégalités (socio-économiques) tandis que la seconde chercherait davantage à éviter que des inégalités n'apparaissent et à garantir l'égalité des chances. On peut

d'ailleurs aussi envisager la justice sociale dans une perspective plus contextuelle, comme le fait MAIEM qui insiste sur la nécessité de considérer différentes perspectives sur la justice, en particulier celles des communautés marginalisées.

PROPOSITIONS DE RECOMMANDATIONS :

La question des biais (déjà évoquée dans la section précédente) et de l'opacité des algorithmes (le « problème de la boîte noire ») a aussi retenu l'attention. Ce n'est pas forcément étonnant dans la mesure où cet enjeu a bénéficié d'une importante couverture médiatique. Parent note par exemple que « les systèmes de prise de décisions assistée, voire automatisée, en médecine, finance, défense ou justice, donneront des résultats biaisés si leurs intrants sont biaisés. » De même, l'OIQ insiste sur la nécessité d'instaurer des « mécanismes de contrôle et de protection » pour corriger les biais.

D'autres recommandations peuvent être mentionnées :

- Il faut former les étudiants et praticiens en IA au droit et à l'éthique. (Parent, OIQ)
- Il faut promouvoir l'emploi diversifié et féminin dans le développement de systèmes d'IA. (OIQ)
- Il faut assurer un traitement rapide et transparent des réclamations des utilisateurs/citoyens qui auraient été affectés par les effets défavorables d'un système d'IA (OIQ).

Plusieurs mémoires appellent de leurs vœux la création d'un **organisme indépendant de supervision** (Parent, McNAlly, Hernandez, OIQ, AQT). Son rôle ne se cantonnerait pas à l'application du principe de justice, mais comme il est souvent mentionné à propos des enjeux de discriminations, on peut saisir l'occasion pour l'évoquer ici.

Sa présentation varie bien sûr selon les mémoires. L'OIQ parle d'un observatoire de l'IA, Hernandez évoque « un organisme régulateur dont la tâche serait d'assurer au citoyen une bonne compréhension des décisions prises par les IA » ; quant à l'AQT, elle préconise « la mise en place d'un comité des sages multisectoriel qui aura pour mission d'engager une démarche réflexive sur les opportunités et les défis de l'industrie québécoise des technologies sur la question de l'éthique en intelligence artificielle. » On peut aussi penser comme le suggère McNally à un organisme de surveillance qui collaborerait étroitement avec le gouvernement et aurait pour mandat d'anticiper les problèmes que l'IA posera à la société de demain.

RESPONSABILITÉ

PRINCIPE PROPOSÉ :

« Les différents acteurs du développement de l'IA devraient assumer leur responsabilité en œuvrant contre les risques de ces innovations technologiques ».

REMARQUES GÉNÉRALES

Le principe de responsabilité est moins mentionné que les précédents dans les mémoires reçus, mais on peut dire que son ombre plane dès que la question de la relation entre les humains et l'IA est soulevée. Qui sera responsable de l'IA, en particulier de ses effets néfastes ? Comme le remarque SRAD, le développement de l'IA pourrait aller jusqu'à l'utilisation de robots tueurs. Cette possibilité soulève à son tour une inquiétude largement partagée : que les humains se déchargent de leurs responsabilités sur le dos de l'IA. On retrouve ici le thème de l'IA comme outil : celle-ci devrait être vue comme une extension de l'intentionnalité humaine, mais non comme une intentionnalité autonome (MAIEM).

Parmi les personnes et groupes responsables, on peut citer les chercheurs qui, détenant la connaissance, doivent soulever le débat (SRAD). Il faut y ajouter la responsabilité de ceux qui commanditent les chercheurs, comme les universités, les militaires ou les industries. Être

responsable signifie notamment mettre en place les savoirs et les outils pour « comprendre le fonctionnement de l'IA et anticiper ses réactions » (MAIEM).

Dans un mémoire qui offre davantage une perspective générale sur la conception de l'IA qui prévaut aujourd'hui qu'une bonification de tel ou tel principe de la Déclaration, Jean-Claude Ravet, le rédacteur en chef de la revue Relations, entend nous mettre en garde contre l'instrumentalisation de l'humain à l'ère de l'IA et estime que le développement de l'IA engage notre responsabilité collective et qu'il est nécessaire d'en avoir une vue d'ensemble, historique et idéologique. Ainsi, c'est le motif même de l'IA comme outil qu'il convient d'interroger, puisque « la frontière entre l'usage de la technique et la technique elle-même se brouille plus que jamais ». Surtout, note Ravet, il convient de ne pas être dupe de l'idéologie qui accompagne le développement de l'IA et sert les intérêts de puissantes multinationales. Pour Ravet, cette idéologie qui entend se faire passer pour un discours scientifique plutôt que pour un projet de société assumé, se caractérise par « une vision extrêmement réductrice de l'humain et de la vie ». (Hernandez s'interroge aussi sur la spécificité de l'humain).

Le mouvement transhumaniste ou le livre *Homo Deus* de Yuval Noah Harari seraient des représentants de cette idéologie réductionniste que condamne Ravet : « Sous prétexte d'augmenter l'humain, on ne doit pas le diminuer et en faire un moyen en vue d'une fin. Le seul critère de rentabilité ne suffit pas. Ni le respect du choix individuel. Car les enjeux touchent au vivant et à l'humanité en tant que tels. » Il importe de jeter un regard critique sur ce qui apparaît bien souvent comme une évidence, à savoir que le progrès de l'humanité passe par l'IA et qu'il y a quelque chose d'inévitable dans le déploiement de ces technologies. Autrement dit, c'est collectivement que nous sommes responsables et c'est bien l'humain qui doit toujours garder le dernier mot, « en tant qu'être de parole, de sentiments, de sensations, et conscient de sa fragile humanité et des liens qui l'unit aux autres, au vivant et à la Terre » (Ravet).

PROPOSITIONS DE RECOMMANDATIONS

> Les décisions de justice faite avec l'assistance d'une IA devront ultimement être imputables à des êtres humains (SRAD, Parent, Ravet).

> Dans le cas des ingénieurs, il faut préserver l'imputabilité des professionnels dans l'exercice de leur profession (OIQ).

> Du point de vue de la responsabilité juridique, il faut anticiper d'éventuels litiges autour des systèmes d'IA avec les juridictions non canadiennes (ex. composantes conçues ou fabriquées ailleurs que là où le système a été utilisé) (OIQ).

> Pour éviter d'attribuer une responsabilité indue aux IA, elles ne devraient pas avoir l'apparence trompeuse d'un patient moral (c'est-à-dire un individu pouvant subir un tort) qui mérite notre empathie (MAIEM).

> Dans la formulation du principe, l'intention de « lutter contre les risques » n'est pas assez exigeante : les responsables devraient assumer les résultats du développement de l'IA (MAIEM).

BIEN-ÊTRE

PRINCIPE PROPOSÉ :

« Le développement de l'IA devrait ultimement viser le bien-être de tous les êtres sentients. »

REMARQUES GÉNÉRALES

Comme la responsabilité, le principe du bien-être est souvent présent de manière implicite, en particulier lorsqu'il est question de santé, de sécurité ou même de juste répartition des bénéfices de l'IA. En fait, selon certaines approches en philosophie morale, c'est même un principe qui pourrait servir de critère général de prise de décision : lorsqu'on

a le choix, on devrait agir de manière à produire le plus de bien-être possible. Évidemment, comme le remarque notamment MAIEM, d'autres valeurs peuvent entrer en conflit avec le bien-être, en particulier l'autonomie. Par exemple, on peut décrire les situations où le paternalisme semble justifié comme des situations où on restreint l'autonomie d'un patient moral au nom de son bien-être. Dès lors, il n'est pas étonnant que la possibilité de tels conflits de valeurs — souvent discutés par les philosophes dans les dilemmes moraux — soit envisagée dans les mémoires reçus. Pour autant, il n'en demeure pas moins vrai qu'un principe portant sur le bien-être se doit d'être simple, facile à comprendre et laisser une certaine marge pour des interprétations futures (MAIEM).

De son côté, l'OIQ remarque que le principe du bien-être s'accorde bien avec une des principales dispositions du code de déontologie de l'ingénieur (article 2.02) qui stipule que « l'ingénieur doit respecter ses obligations envers l'homme et tenir compte des conséquences de l'exécution de ses travaux sur l'environnement et sur la vie, la santé et la propriété de toute personne. » C'est pourquoi promouvoir le bien-être suppose d'évaluer le plus finement possible les risques liés au déploiement et au fonctionnement des applications de l'IA, en gardant en tête que « le risque zéro n'existe pas » (OIQ).

Il faut aussi noter que le caractère très inclusif de ce principe, qui vise non seulement le bien-être des humains, mais aussi de l'ensemble des êtres sentients n'a pas été remis en question. C'est peut-être le signe d'une évolution des mentalités quant à notre rapport aux animaux non humains (sentients). MAIEM et SRAD reprennent à leur compte cette extension du domaine de la moralité aux êtres sentients tandis que Parent évoque les interférences de l'IA avec la vie animale. On peut aussi constater que certains mémoires (Ravet, MAIEM, Parent) semblent vouloir tenir compte du critère de la vie pour étendre le cercle de la moralité à des entités non sentientes (comme des végétaux ou des écosystèmes). Ces intuitions qu'on pourrait qualifier de biocentristes ne sont toutefois pas assez développées pour qu'on puisse dire qu'elles reflètent une position morale (assez radicale) assumée : il peut s'agir d'un souci pour l'environnement d'ordre anthropocentriste.

L'idée semble aussi partagée que la capacité pour une IA d'être sentiente (ou sensible) serait un bon critère pour lui conférer des droits ou, à tout le moins, de la considération morale puisque si un robot pouvait souffrir, par exemple, il aurait alors un intérêt légitime à ce qu'on le protège. Ce point demeure toutefois très spéculatif dans la mesure où les systèmes d'IA sont actuellement très loin d'éprouver des sensations ou des émotions.

Enfin, dans un texte assez spéculatif et programmatique, Museau cherche à articuler la notion de minimalisme morale développée par le philosophe Ruwen Ogien et le principe du bien-être proposé par la Déclaration. Il en ressort que le développement de l'IA devrait viser à ne pas faire de tort à autrui et non pas à s'améliorer soi-même — l'amélioration de soi étant, selon Museau, à la fois de l'ordre du maximalisme moral et du transhumanisme.

PROPOSITIONS DE RECOMMANDATIONS

> Formuler le principe en termes de réduction des souffrances plutôt que de promotion du bien-être (ce qui correspond à ce qu'on nomme parfois l'utilitarisme négatif) (MAIEM).

> Par souci de sécurité, il faut prévoir des dispositifs de débrayage/blocage dès la conception des systèmes d'IA afin d'en conserver le contrôle en cas de défaillance (OIQ).

AUTONOMIE

PRINCIPE PROPOSÉ :

« Le développement de l'IA devrait favoriser l'autonomie de tous les êtres humains et contrôler, de manière responsable, celle des systèmes informatiques ».

REMARQUES GÉNÉRALES

Concernant l'autonomie, on peut dire qu'il y a un consensus quant à promouvoir l'autonomie des humains. Cette idée se traduit notamment par le thème déjà évoqué de l'IA *au service* des humains. L'OIQ note ainsi que « les robots et les systèmes d'IA doivent être vus comme des outils d'assistance ou d'aide à la décision et non comme une substitution au jugement humain ». De son côté, Ravet insiste pour ne pas réduire l'humain à une machine ni en faire un moyen en vue d'une fin, tandis qu'Hernandez se demande si l'IA ne va pas un jour remplacer les humains de sorte qu'ils deviennent obsolètes.

Il n'en demeure pas moins vrai que la notion d'autonomie renvoie à de multiples acceptions. SRAD propose ainsi une grille d'analyse détaillée des types d'autonomie (« condition d'une entité qui choisit elle-même les lois auxquelles elle se soumet ») divisée en autonomie morale, politique et fonctionnelle (non-dépendance). Ces trois types d'autonomie peuvent être croisés avec trois types de situations : l'autonomie d'un humain aidé par une IA (par exemple une personne en situation de handicap), l'autonomie d'un humain dans un environnent peuplé d'IA et, enfin, l'autonomie d'une IA dans un environnement humain. SRAD propose dès lors une reformulation qui tienne davantage compte de ces diverses acceptions : « Les systèmes IA ne doivent pas nuire à l'autonomie (morale, fonctionnelle, politique) des êtres humains, mais devraient chercher à y contribuer. Les systèmes IA ne doivent pas être rendus entièrement autonomes des êtres humains, mais doivent demeurer sous leur contrôle (moral, fonctionnel, politique) ». Pour autant, on n'en déduira pas trop vite que l'autonomie devrait systématiquement prévaloir sur les autres valeurs comme le bien-être, la justice ou la connaissance. Chaque cas est à examiner en contexte. Et comme le rappelle MAIEM, le consentement des gens demeure une bonne manière de garantir leur autonomie.

Si la valeur de l'autonomie humaine fait consensus, la situation est plus délicate pour ce qui a trait à « l'autonomie des systèmes d'IA » dans la mesure où leur mise sous tutelle pourrait être contestée. Ainsi, citant un article sur l'évolution digitale et la vie artificielle, MAIEM rappelle qu'on pourrait imaginer des situations dans lesquelles l'autonomie et la créativité des systèmes d'IA seraient profitables au bien-être général. Il n'empêche, poursuit MAEIM, que l'autonomie d'un système d'IA ne devrait pas être recherchée pour elle-même si cela entre en conflit avec le bien-être d'un être sentient. Pour pertinentes qu'elles soient, ces remarques sont assez isolées parmi les mémoires : ceux-ci donnent plutôt le sentiment qu'il faut étroitement surveiller les systèmes d'IA au risque d'en perdre le contrôle. Il semble toutefois envisageable de concilier ces considérations apparemment divergentes : on pourrait garder le contrôle à un certain niveau sur un système d'IA tout en autorisant — à un plus bas niveau et dans un cadre déterminé — que l'IA expérimente certaines solutions à des problèmes de façon libre et créative.

PROPOSITIONS DE RECOMMANDATIONS

> La notion d'autonomie a davantage suscité des réflexions philosophiques que des recommandations concrètes, même si certaines recommandations des autres sections ne sont pas sans rapport (par exemple celles sur le consentement dans la section vie privée).

CONNAISSANCE

PRINCIPE PROPOSÉ :

« Le développement de l'IA devrait promouvoir la pensée critique et nous prémunir contre la propagande et la manipulation ».

REMARQUES GÉNÉRALES

Les liens entre l'IA et la connaissance sont multiples. Tout d'abord, et dans la perspective des sciences cognitives, l'intelligence artificielle peut nous aider à comprendre l'intelligence naturelle, l'une et l'autre pouvant se définir comme ce qui guide une capacité d'action (SRAD). Dès lors, on peut se demander pourquoi l'intelligence naturelle devrait prévaloir sur l'intelligence artificielle puisqu'à un certain niveau d'analyse, les humains ou les animaux, tout comme les machines, sont des systèmes causaux.

Dans plusieurs mémoires, le principe de connaissance est l'occasion de discuter des enjeux de propagande et des fausses informations (*fake news*). De ce point de vue, le sujet concerne tout autant la démocratie que la connaissance. Ainsi, on peut se demander comment une IA ou ceux qui la produisent et la commercialisent pourraient être en position de décider ce qui relève de la propagande ou de la manipulation. Il semble illégitime, voire dangereux de leur confier une telle responsabilité. C'est pourquoi MAIEM propose une reformulation du principe qui mette plutôt l'accent sur la transparence : « Le développement de l'IA ne devrait pas nuire à la pensée critique. Il devrait également être transparent et ouvert afin de permettre la participation du public au développement de l'IA, un contrôle public de l'IA et l'éducation à l'IA. » [*The development of AI should not hamper critical thinking. It must also proceed in a transparent and open manner, to enable public participation in its development, scrutiny, and education*].

Parmi les autres thèmes liés à la connaissance, on trouve l'accès public aux résultats des recherches en IA, la pensée critique (MAIEM nous met en garde contre les chambres d'écho), l'éducation à l'IA et l'opacité des algorithmes déjà mentionnés dans la section justice. Sur ce dernier point, SRAD en appelle à des efforts pour améliorer la transparence des données et des algorithmes, mais aussi pour publier les codes sources derrière les IA.

PROPOSITIONS DE RECOMMANDATIONS

> Des mesures devraient être mises en place afin de promouvoir l'accès public aux résultats des recherches universitaires. (MAIEM)

> Il faut encourager la compétition et la diversité dans les applications de l'IA afin que cela bénéficie à toute la société. (MAIEM)

> Il faut repenser le *business* modèle des médias sociaux des autres sites de nouvelles. (MAIEM)

> Tous les étudiants et praticiens en IA devraient recevoir une formation avancée en éthique. (Parent)

DÉMOCRATIE

PRINCIPE PROPOSÉ :

« Le développement de l'IA devrait favoriser la participation éclairée à la vie publique, la coopération et le débat démocratique »

REMARQUES GÉNÉRALES

Concernant la démocratie, plusieurs mémoires (Robert, Parent, OIQ, AQT, SRAD) saluent l'initiative de la Déclaration et la possibilité qu'elle leur offre

de faire entendre leur point de vue. MAIEM y voit une « contribution importante » aux discussions internationales sur le sujet.

D'autres sont plus critiques. Ainsi, Quintal et al. contestent le processus même de production de la Déclaration de Montréal. S'ils sont favorables aux efforts de consultation publique, ils se demandent toutefois s'il ne s'agit pas là d'une entreprise de légitimation visant à entériner un document existant. Plus précisément, ils craignent que la version préliminaire de la Déclaration (soit les 7 principes proposés qui servent d'articulation à cette synthèse) n'ait fortement orienté les débats citoyens : « Le public aurait dû être impliqué dans la délibération sur le contenu de la Déclaration dès le tout début » [the public should have been meaningfully engaged in deliberating on the contents of the Declaration from the very beginning]. Pour Quintal et al., cela risque de compromettre la légitimité finale de la Déclaration.

Ces inquiétudes signifient évidemment un appel à plus de démocratie (et de transparence et d'esprit critique) dans le développement de l'IA, ce qui appuie en quelque sorte le principe lié à la démocratie. En outre, précisent Quintal et al., la bonne volonté démocratique restera un vœu pieux si elle ne s'accompagne pas d'une régulation de l'industrie. On court aussi le risque que les compagnies utilisent les algorithmes pour confiner le débat à des enjeux qui leur paraissent acceptables (ce qu'on pourrait avec SRAD qualifier d'enjeu épistémique ayant des effets néfastes pour la démocratie). Un argument similaire est proposé par MAIEM qui note que puisqu'il est peu probable que les compagnies partagent leurs algorithmes afin de protéger leur propriété intellectuelle, la régulation externe semble être la meilleure solution.

Sur le principe lui-même, MAIEM trouve sa formulation un peu vague et regrette qu'il se concentre sur la démocratie alors que tous les humains ne vivent pas dans de tels régimes. MAIEM suggère dès lors de lui substituer un « principe de participation publique » qui se présenterait ainsi : « Le développement de l'IA devrait s'accompagner d'une information claire et précise afin de permettre un débat éclairé sur l'IA et ses applications, tout en encourageant l'ouverture et la transparence dans la recherche. » [*The development of AI should promote the dissemination of clear and accurate information to the public to enable open and educated debate about AI and its applications, and encourage open and transparent research collaboration.*]

SRAD, enfin, rappelle que les grandes entreprises de technologie (comme les GAFA) possèdent aujourd'hui un pouvoir considérable, tant économique que politique — notamment parce qu'elles ont un accès direct à énormément de données personnelles. On peut y voir une menace sérieuse pour la démocratie comme le suggère l'affaire Cambridge Analytica qui a éclaté depuis lors. Par ailleurs, dans la mesure où la démocratie requiert une certaine égalité socioéconomique — au risque de dégénérer en oligarchie — il faut prendre garde à l'accroissement des inégalités que devrait mécaniquement entrainer le développement de l'IA. En effet, explique SRAD, l'automatisation d'une tâche par une IA revient à déplacer la richesse du revenu vers le capital (et donc à la concentrer entre les mains des actionnaires plutôt que des salariés remplacés par l'IA). À moins de régulations ou d'encadrement, l'IA risque donc d'amplifier la croissance des inégalités économiques qu'on constate depuis les années 1950.

PROPOSITIONS DE RECOMMANDATIONS :

> Les chercheurs à l'avant-garde du secteur dans nos institutions publiques universitaires doivent garder leur indépendance face à l'entreprise privée. (Parent)

> Il faut briser les grands monopoles dans l'industrie technologique. (SRAD)

> Il faut sérieusement considérer la possibilité d'un revenu universel garanti financé par une taxe sur l'automatisation ou le capital. (SRAD)

> Il faut encourager de nouvelles structures de propriété d'entreprises telles que des coopératives pour lutter contre la concentration du capital. (SRAD)

< >
Déclaration de Montréal
IA responsable_
</ >

PARTIE 6

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

CRÉDITS

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

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 vieillis 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 créative, 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 accorde[3]. 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 algorithmes[1] ?

2.1

COMMENT GOUVERNER LES ALGORITHMES : PROMOUVOIR LA PARTICIPATION CITOYENNE

Selon Antoinette Rouvroy et Thomas Berns, la gouvernance algorithmique se déploie en trois temps[4] :

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 connaissances[5]. 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 européenne[6], 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 « gouvernementalité algorithmique »

[5] Antoinette Rouvroy et Thomas Berns. 2013. « Gouvernementalité 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/

conditions[7]. La Ville de Montréal développe des politiques sur l'ouverture des données[8] et sur le logiciel libre[9] 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 décideurs[10]. 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 législatif[12]. 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 genres[13].

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 SIA[14]. 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 emboité 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 sociaux[17]. 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 SIA[19]. 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 SIA[20].

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, déplait 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 oligopolistique[21]. 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 petits[22].

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 défense[23]. 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 algorithmes[24]. Cela pose le défi de la coopération internationale et des rivalités entre États qui veulent asseoir leur hégémonie[25] 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-qBZUXEUG4dVk909eSg5qvfbpNlRhzlefWPtBwbxY/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/GovAIAgenda.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.

multilatérales[26]. 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 interrogations[27] : 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 algorithmique[30]. Si la participation et la collaboration des parties prenantes requièrent du temps, elles n'en sont pas moins sans valeur[31]. 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'omniscience[32].

[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 P*hilosophical 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 public[33]. 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 donnée[34]) 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'expression[36]. Cette organisation dépend des formes de régulation décidées par les acteurs eux-mêmes[27]. 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 compagnies[37]. 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. Section 6 de ce rapport, 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 publiques.

[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 imprévues[38]. Le danger réside ici dans l'omnipotence du langage informatique qui façonne le monde des possibles sans égard aux subtilités inhérentes au contexte social[39]. 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 consommateurs[40]. Les « profils numériques » sont ainsi utilisés, parfois à l'insu des individus, pour différentes fins, au risque de se substituer à leur identité propre[28]. 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 désubjectivation est alors souligné[41]. 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 individus[42]. 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 humains[43]. 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 algorithmes[44], 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 individus[45], 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 réfèrer 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 produisent[37]. 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 bénéfices – 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 tuer[48]).

Bien qu'à court et moyen terme, l'humain semble destiné à conserver le contrôle sur l'IA[49], 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 droit[50]. 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 connectés[51]. 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 biomédicales[52]. 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'impose[53]. Certains philosophes introduisent l'idée de droit à l'intériorité[54] alors que des informaticiens expérimentent, avec des résultats mitigés[55], des techniques d'anonymisation des données personnelles afin de prévenir la (ré)identification.

Pour bon de nombre de chercheurs, l'opacité des réseaux neuronaux constitue précisément le nœud du problème[56]. Et dans le secteur public, l'enjeu est de taille car les algorithmes prennent des décisions qui ont un impact majeur sur la vie quotidienne[57]. Sans explication, surtout en cas d'erreurs ou de dysfonctionnements, et sans recours, les préjudices commis pourraient pénaliser injustement les individus[58], 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 dissimulées[59]. 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 neutres[60]. 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 affectés[61] (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 citoyens[62]. 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 Ruijer 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 capabilités[63] 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 capabilités 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 capabilités 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 objectifs[64].

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 capabilités 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 capabilité 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 médiatique[66]. 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 humain[68] 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 critique[69] :

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 secondaire[70]. Le programme finance également la formation et le perfectionnement professionnel des nouveaux enseignants par l'intermédiaire d'HabiloMédias qui crée plusieurs ressources en ligne[71]. 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 séniors, 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 autres[73]. 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 vie[75].

[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-pédagogiques

[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 géographie[76]. 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 REPTIC[78] 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 supérieure[79]. 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'entrainement 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-pédagogiques/résultats-dapprentissage-en-éducation-aux-médias-et-littératie-numérique-par-province-et-territoire/québec-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.uquebec.ca/pdci/referentiel-de-competences-informationnelles-en-enseignement-superieur

[80] Québec: Ministère Travail, Emploi et Solidarité sociale. 2018. « Stratégie nationale sur la main d'oeuvre 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 « accroitra 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 Canada[85] 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 informatique[86]. Mieux formés aux enjeux éthiques et sociaux des SIA et des système 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 potentiels[87].

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 humaniste[88] 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 société[90] 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 ». 9e *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 dépendance[93], 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 numériques[94]. 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 québécois[97].

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 territoire[98]. 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-premire-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.pdfCEST-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 labs[99], 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 insidieuse[100]. 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 émotionnel[104], 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 erronées[105]. 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 escompté[106]. 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-né[107]. 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 formulent 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 corréler 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 discriminations[108].

Les récents travaux de Virginia Eubanks[109] 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 Cheklist 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 marginalisation[110]. Ces outils de prédiction, par boucle de rétroaction, sont susceptibles de créer ainsi les difficultés qu'elles prétendent signaler[111]. 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'Neil[112]. 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-américaine[114]. 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 blancs[115].

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 utilisé[116]. 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 Biais ». 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 réfugiés[117]. 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'adressent[118]. 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 ciblées[119]. 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 utilisateurs[123] 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 « Institutionality and algorithms », 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'inconnu[124]. 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 canalisaient 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 groupes[125]. 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 sexuels[129]. Cela a comme effet de renforcer les stéréotypes culturels[130] et de dissuader les gens de faire des recherches impopulaires[131].

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.

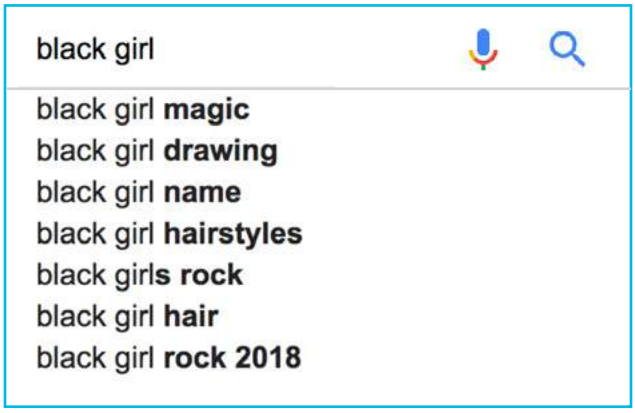
[131] L. Gannes. 2013. « Nearly a Decade Later, the Autocomplete Origin Story: Kevin Gibbs and Google Suggest ». *All Things D. Accessed* January 29: 2014.

[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'entrainement).

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 interprétées[133].

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 voir[134]. 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'entrainement, les biais humains comme les présomptions culturelles sont aussi transmis à travers les choix de classifications[135]. 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 données[136].

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 Revie*w 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'entrainement, et inversement qu'une autre population est sous-représentée[137]. Ainsi, les données sur lesquelles s'entraine un algorithme peuvent être biaisées ou faussées et présenter un échantillon non représentatif qui a été pauvrement défini avant usage[138]. Un bon exemple est celui des SIA de reconnaissance faciale : plus il y a de visages de personne blanche dans les données d'entrainement, plus le système sera performant pour cette partie de la population[139]. 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 gorille[140].

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 probantes[141]. 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 défaillantes 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 SÉRENPIDITÉ (LA FORTUITÉ)

Le terme de sérendipité vient de l'anglais *serendipity*, un mot qui a été formé par l'écrivain britannique Horace Walpole en 1754[143]. 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 importante[144] 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 #hastag pour Twitter, Chris Messina[145].

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 intellectuelle[146], ce que l'on peut admettre[147]. Mais l'idée de *black box* a une autre connotation : on soupçonne que les compagnies ne veuillent 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 privées[148]. 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 appeler 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 spécialistes[149], 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 service[150].

Représentativité et inclusivité

Pour assurer une IA inclusive, on ne doit pas seulement s'intéresser à la conception et l'entrainement 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 similaire[151]. On peut penser que leurs intérêts et leurs expériences de vie influencent leur conception des algorithmes et leur programmation[152]. 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 emplois[155] ? 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 artificiel[156].

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 Unies[158] 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 complet[160].

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 écologique[161]. 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 XXIe 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 évènements 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'être[164].

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'usagers de la 4G dans le monde d'ici 2025, et de 1,2 milliard de nouveaux usagers de la 5G[165].

[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 2025[166], 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 Cisco[167], 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 Project[168] 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 production[169]. 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 mainsstream 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 Schift Project (2018). Idem que la précédente.

consommateurs d'énergie[170]. 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 difficiles[171].

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 mondiale[173].

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 maitrisée 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 $GtCO_2eq$: une mesure d'émissions de gaz à effet de serre

Il existe plusieurs gaz à effet de serre. Si le dioxyde de carbone, ou CO_2, est responsable de 76 % du réchauffement planétaire dû à l'activité humaine, d'autres doivent aussi être considérés, comme le méthane CH_4 ou le protoxyde d'azote N_2O[174]. Chaque gaz a un potentiel de réchauffement global (PRG) différent. Le CO_2 est considéré comme référence : son PRG est 1. Le méthane, par exemple, a un PRG de 25 : une tonne de CH_4 a donc un impact 25 fois plus grand qu'une tonne de CO_2. Le PRG permet de comparer les émissions de gaz différents, en utilisant la tonne de CO_2 équivalente (tCO_2eq) comme unité.

En 2016, le Canada a produit 704 $MtCO_2eq$[175], soit l'équivalent de 704 millions de tonnes de CO_2. La même année, le monde produisait autour de 50 $GtCO_2eq$.

[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 rebonds[176]. 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'accentue dans les récentes publications de l'UNEP[177] et du GIEC[178], 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 $GtCO_2eq$ par an, devraient être réduites de 10 $GtCO_2eq$ d'ici 2030 pour avoir une chance de tenir l'objectif de 2 °C, et de 20 $GtCO_2eq$ 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 CO_2eq 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 artificiel[179].

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 (Environnemental 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 sensibilisation[181]. 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'équipements[182], 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 services[183], 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 internet[185].

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'ISO[186] 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 résulteront ? 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 conséquentielle 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 Now[187] 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 réflexion[188]. 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 déployées : 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 chaine 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 PRÉDICTIFS 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 labelisées « 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 évènements 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 évènements climatiques extrêmes et au réchauffement (nouveaux ravageurs) qui peuvent entrainer 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 multiples[189]. 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 entrainer un certain « solutionisme » (résolution locale d'un problème grâce à un outil maitrisé, 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 grid*s 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 entraine-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 littératie 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.

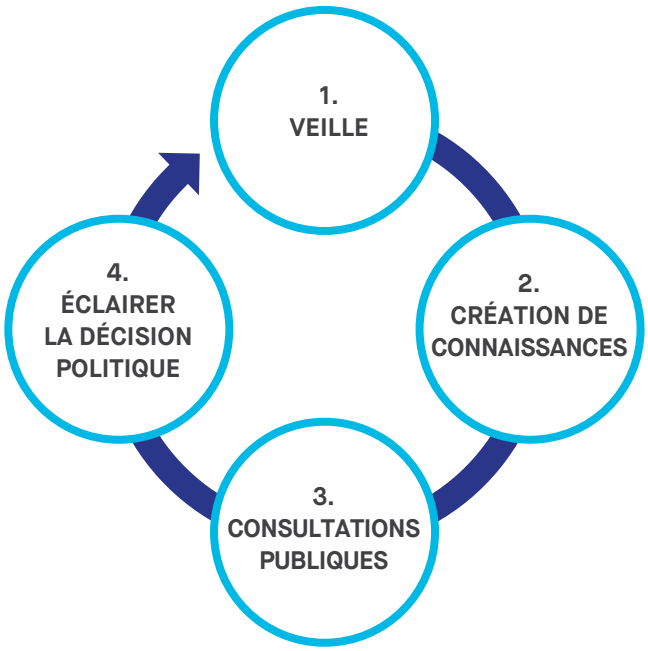
6. LES RECOMMANDATIONS EN VUE DE L'ÉLABORATION DE POLITIQUES PUBLIQUES

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'usagers.

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, boites 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 partis concernés 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.

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 :

5.1.1 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'entrainement 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'entrainement utilisées par les SIA publics comprennent un échantillon représentatif des populations concernées.

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 accroitre 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

CRE
CENTRE DE RECHERCHE EN ETHIQUE

LAB VILLE PROSPECTIVE

IVADO
INSTITUT DE VALORISATION DES DONNÉES

SAT

Mila

ICRA
Programme
IA et
société

Université de Montréal | designNsociété

polEtHics
CHAIRE DE RECHERCHE DU CANADA
ÉTHIQUE PUBLIQUE ET THÉORIE POLITIQUE

CIRANO
Allier savoir et décision
Centre interuniversitaire de recherche en analyse des organisations

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

UNIVERSITÉ LAVAL
Institut d'éthique appliquée

MUSÉE DE LA CIVILISATION
Québec

CENTRE D'ÉTUDES ET DE RECHERCHES INTERNATIONALES
CÉRIUM
Université de Montréal

crdm_ul
CENTRE DE RECHERCHE EN DONNÉES MASSIVES DE L'UNIVERSITÉ LAVAL

Canada

Centre Culturel Canadien Paris
Canadian Cultural Centre Paris

MEC
Maison des étudiants canadiens

< >

Déclaration de Montréal
IA responsable_

</ >

declarationmontreal-iaresponsable.com