



# BILAN DES DÉLIBÉRATIONS CITOYENNES DE LA DÉCLARATION DE MONTRÉAL IA RESPONSABLE

JUIN 2018



# TABLE DES MATIÈRES

CRÉDITS	6	4.2 La méthodologie des ateliers de coconstruction	36
SOMMAIRE	7	4.3 Originalité de la démarche de coconstruction	38
INTRODUCTION	10	4.4 Cafés citoyens en marge des bibliothèques	39
		4.5 Portrait des participants	40
<b>1. POURQUOI LA DÉCLARATION DE MONTRÉAL IA RESPONSABLE?</b>	<b>13</b>	<b>5. PARCOURS DÉLIBÉRATIFS DANS LES ATELIERS : EXEMPLES DE DEUX SECTEURS : VILLE INTELLIGENTE ET MONDE DU TRAVAIL</b>	<b>43</b>
1.1 L'origine intellectuelle du projet	14	5.1 Les parcours délibératifs	44
1.2 Le Forum sur le développement socialement responsable de l'intelligence artificielle	16	5.1.1 Secteur ville intelligente : la voiture autonome (VA) et le juste partage de la rue	45
1.3 Vers la <i>Déclaration de Montréal</i> pour un développement responsable de l'IA	17	5.1.2 Secteur du monde du travail : une restructuration socialement responsable?	49
1.4 Montréal et le contexte international	18		
<b>2. LA VERSION PRÉLIMINAIRE DE LA DÉCLARATION DE MONTRÉAL POUR UN DÉVELOPPEMENT RESPONSABLE DE L'IA</b>	<b>21</b>	<b>6. LES PREMIERS RÉSULTATS DE LA COCONSTRUCTION</b>	<b>54</b>
Préambule	22	6.1 Résumé	55
Valeurs et principes	23	6.2 Les données de la coconstruction : notes explicatives	56
		6.3 Développement responsable de l'IA : les grandes directions attendues par les citoyens	57
<b>3. LES ENJEUX ÉTHIQUES ET SOCIÉTAUX DE L'IA</b>	<b>24</b>	6.4 Résultats : la perception citoyenne des enjeux	59
3.1 Se faire une idée de l'IA	25	6.4.1 Les principaux enjeux abordés par secteur	60
3.2 L'IA au quotidien et le questionnement philosophique	27	6.4.2 Des enjeux qui pourraient conduire à la création de nouveaux principes	63
3.3 Les enjeux éthiques de l'IA	29	6.4.3 Des enjeux relatifs au rapport entre humains et IA	65
3.4 L'éthique de l'IA et la Déclaration de Montréal	31	6.5 Résultats : les pistes de solution	66
		6.5.1 Les grandes pistes de solutions proposées par les citoyens	66
<b>4. LA DÉMARCHE DE COCONSTRUCTION</b>	<b>32</b>	6.5.2 Pistes de solution par secteur	69
4.1 Les principes de la démarche de coconstruction	33		
4.1.1 Les principes d'une bonne participation citoyenne	33		
4.1.2 Experts et citoyens	35		

# TABLE DES MATIÈRES

7.	UN PROCESSUS DE COCONSTRUCTION EN CONTINU	77	Tableau 1 : Ville intelligente, Premier moment délibératif: formulation d'enjeux éthiques en 2025	46
7.1	Poursuivre la délibération	78	Tableau 2 : Ville intelligente, Deuxième moment délibératif: propositions d'encadrement de l'IA pour 2018-2020	47
7.2	Une démarche inspirante	79	Tableau 3: Monde du travail, Premier moment délibératif: formulation d'enjeux éthiques en 2025	51
<b>CONCLUSION</b>		<b>80</b>	Tableau 4 : Monde du travail, Deuxième moment délibératif: proposition d'encadrement de l'IA pour 2018-2020	52
PARTICIPANTS À LA COCONSTRUCTION		83	Tableau 5 : Les principes auxquels réfèrent les enjeux prioritaires identifiés par les citoyens	55
<b>ANNEXES</b>		<b>85</b>	Tableau 6 : Les pistes de solution proposées pour répondre aux enjeux identifiés	56
ANNEXE 1 – Les ateliers de coconstruction : description détaillée et fonctionnement		86	Tableau 7 : Enjeux prioritaires identifiés par les citoyens en fonction des principes de la Déclaration (nombre de tables)	59
	Les cafés citoyens	86	Tableau 8 : Nombre de tables qui proposent chacune des catégories de pistes de solution	68
	Les journées de coconstruction	87	Tableau 9 : Pistes de solution ou grandes directions pour le secteur de l'éducation	69
ANNEXE 2 – Les scénarios prospectifs		88	Tableau 10 : Pistes de solution ou grandes directions pour le secteur du système judiciaire et de la police prédictive	71
ANNEXE 3 – les autres formes de participation		96	Tableau 11 : Pistes de solution ou grandes directions pour le secteur de la santé	73
	A3.1 Le questionnaire en ligne	96	Tableau 12 : Pistes de solution ou grandes directions pour le secteur monde du travail	74
	A3.2 Les mémoires ou recommandations spontanées	97	Tableau 13 : Pistes de solution ou grandes directions pour le secteur de la ville intelligente et des objets connectés	75
<b>TABLE DES TABLEAUX ET DES FIGURES</b>			Tableau 14 : Déroulement type des cafés citoyens	86
	Figure 1 : Les valeurs de la Déclaration	14	Tableau 15 : Déroulement type des journées de coconstruction	87
	Figure 2 : La démarche de coconstruction	17	Tableau 16 : Résumé des scénarios	88
	Figure 3 : La prospective stratégique : une démarche en trois temps	38	Tableau 17 : Constitution de cinq scénarios par thème	91
	Figure 4 : proportion hommes-femmes ayant participé aux ateliers de coconstruction	40		
	Figure 5 : Les participants aux ateliers de coconstruction par tranches d'âge	40		
	Figure 6 : Répartition des répondants aux cafés citoyens et aux journées de coconstruction par niveau de scolarité atteint	41		
	Figure 7 : Répartition des répondants aux cafés citoyens et aux journées de coconstruction par secteur d'activité	42		
	Figure 8 : Trois grandes pistes de solutions prioritaires à toutes les tables	66		

# NOS PARTENAIRES

Université   
de Montréal



CENTRE DE RECHERCHE EN ETHIQUE



**ICRA**  
Programme  
IA et  
société



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



# CRÉDITS

Ce rapport sur les résultats des délibérations citoyennes a été rédigé sous la direction de

**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)

**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, titulaire de la chaire de recherche du Canada en Éthique publique et théorie politique

## Coordination scientifique

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

Assistée de

**Camille Vézy**, doctorante en communication, Université de Montréal

## Équipe de rédaction

**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

La rédaction de ce rapport n'aurait pu être possible sans la contribution des étudiants et professionnels qui ont animé les ateliers et qui ont réalisé la prise de notes, ni sans l'implication de Marie D. Martel, professeure à l'École de bibliothéconomie et sciences de l'information de l'Université de Montréal.

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

Tous les noms se trouvent en conclusion de ce rapport.

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

**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

**Nathalie 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

## Équipe de révision et d'édition

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

**Chantal Berthiaume**, gestionnaire de contenu et rédactrice

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

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

## Travaux coordonnés par

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

# 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*). Huit mois plus tard, nous présentons les premiers résultats du processus de délibération citoyenne qui est à mi-parcours. Ce bilan est très fructueux : 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 collectif et individuel, 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 valeurs fondamentales : l'autonomie, la justice, le bien-être, la vie privée, la démocratie, la connaissance et la responsabilité.

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 des informations en fonction de catégories, 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 les valeurs et les principes de 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 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.

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 de faire cadrer les intérêts publics avec les intérêts privés et d'empêcher l'apparition de monopoles, voire de limiter l'influence des entreprises par des mesures plus coercitives et légalistes.

Les participants recommandent également de mettre en place des mécanismes qui émaneraient et impliqueraient des personnes formées et indépendantes, afin de favoriser la diversité et l'intégration des plus vulnérables, et de protéger la mixité des modes de vie.

Quelles que soient ses 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.

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

Le principe de responsabilité a été jugé l'enjeu le plus souvent prioritaire, suivi de celui d'autonomie, 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.

Pour ce qui est du principe d'autonomie, très souvent choisi comme prioritaire, il a trait à la préservation, voire à 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.

Le principe de bien-être occupe également une place importante pour les participants. Il est présent en filigrane dans les discussions 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 pouvant conduire à la création de nouveaux principes, ou de nouveaux thèmes à explorer à délibérer.

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

Le principe de justice a été abordé selon deux types d'enjeux, ce qui pourrait donner lieu à 2 nouveaux principes : un **principe de diversité** visant à éviter les discriminations en trouvant des mécanismes dépourvus de biais, et 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 de transparence** : 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.

Quel que soit le secteur, les citoyens ont identifié de nombreux enjeux relatifs au rapport entre êtres humains et IA.

### Les grandes pistes de solution

Les participants de toutes les tables de coconstruction se sont entendus 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;
2. La mise en place de formations pour tous et
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 une première phase sur 5 secteurs clés : éducation, santé, travail, ville intelligente et police prédictive et système judiciaire. Une année complète de coconstruction ne suffirait pas à couvrir tous les thèmes de réflexion. La démarche de la coconstruction se poursuivra donc à partir de septembre 2018, permettant d'aborder de nouveaux thèmes qui n'ont été qu'effleurés dans les scénarios utilisés lors de la première phase de coconstruction. Parmi ceux-ci : environnement, démocratie, propagande et médias, ainsi que sécurité et intégrité.

C'est autour de chantiers prioritaires que nous présenterons les recommandations de politiques publiques. À ce jour, nous pouvons affirmer que 3 chantiers se sont imposés : littératie numérique, diversité et inclusion, et transition et mutations sociales. Les résultats finaux seront présentés en décembre 2018.

# INTRODUCTION

## 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*).

Nous n'anticipions pas l'intérêt qu'allait susciter cette démarche, ni l'ampleur de la tâche qui nous attendait. Huit mois plus tard, nous présentons les premiers résultats du processus de délibération citoyenne qui est à mi-parcours. Ce bilan est très fructueux : 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 professionnels.

Le rapport de mi-parcours que nous présentons doit se lire comme un bilan provisoire et non exhaustif d'une expérience de délibération démocratique pour éclairer les choix de politique publique en matière d'intelligence artificielle. 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 des 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 croyons 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, c'est-à-dire qu'il s'agit d'anticiper des controverses éthiques sur des technologies d'intelligence artificielle en devenir

ou sur des situations sociales où l'usage de ces technologies est poussé à la limite de ce que nous pouvons anticiper. Ensuite, nous avons conduit cette démarche de consultation avec une ampleur inédite. Les chiffres cités plus haut l'indiquent clairement. Cette démarche, disons-le dès maintenant, va se poursuivre, et comme la *Déclaration de Montréal* restera un objet ouvert à la révision, la coconstruction ne prendra pas fin au terme de ce premier parcours délibératif.

Nous avons interpellé le public autour de la rédaction de la *Déclaration* et nous avons été interpellés : 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 pour une plus grande transparence de l'intelligence artificielle, nous sommes aussi engagés pour plus de transparence de l'intelligence humaine et collective. Ce rapport de mi-parcours, 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'IA 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.

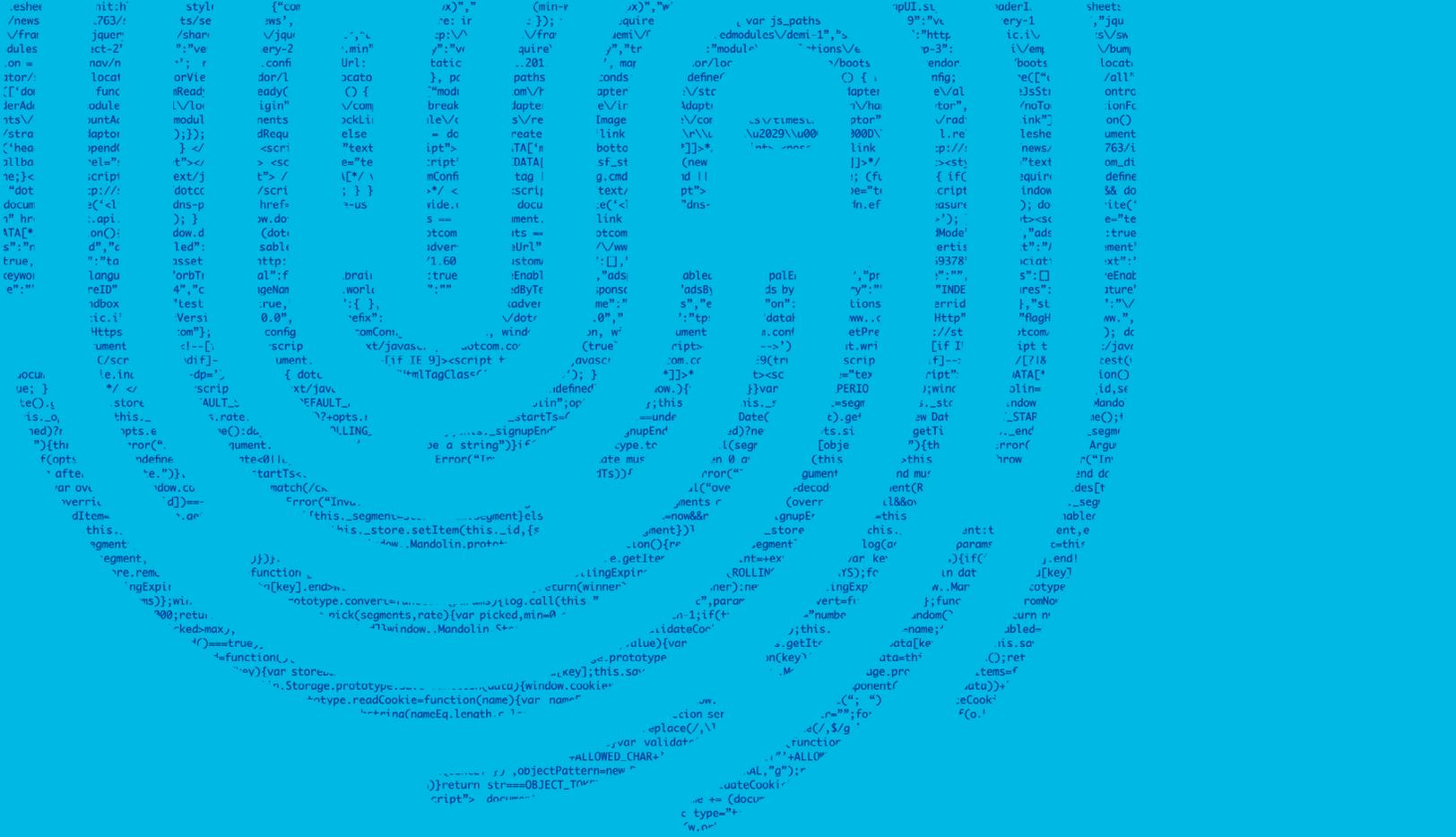
De nombreuses objections et craintes se sont exprimées lors de ce premier parcours de coconstruction. Plusieurs participants aux ateliers et plusieurs observateurs des travaux de la Déclaration ont mis en cause l'idéologie techniciste qui voit dans la technologie le moyen d'organiser rationnellement 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 du revers de la 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. On peut déplorer les effets des nouvelles technologies de l'information et de l'intelligence artificielle 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 ou de préserver des intérêts éthiques et sociétaux fondamentaux.

Pour finir, nous ne trancherons pas la question lancinante concernant l'usage du terme « intelligence artificielle » : est-il approprié pour désigner les algorithmes de traitement d'informations, de reconnaissance et de décision? On peut en contester l'usage en opposant le fait que l'intelligence artificielle renvoie à des processus de connaissance très limités en comparaison avec l'intelligence humaine ou l'intelligence du comportement des pigeons. C'est indéniable. Mais à ce rythme, 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 l'intelligence tout court. Qu'est-ce que l'intelligence humaine? Les centaines de milliers de pages qui ont été produites pour répondre n'y suffisent pas.

Cependant, quelques remarques peuvent permettre d'éviter certains malentendus au fondement même de cette controverse : tout d'abord, on confond souvent l'intelligence et la pensée. L'intelligence est une propriété de la pensée, elle n'est pas toute la pensée. Ensuite, 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 appauvrit la réalité pour mieux agir dessus. 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. Parfois, les procédures sont inadéquates, l'analyse mauvaise. Mais pour l'établir, il faut encore analyser et utiliser des procédures.

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 y a quelque chose de tragique à vouloir éviter la réflexion sur les finalités morales en se contentant d'un calcul sur les moyens. Ceci dit, il est vrai qu'une intelligence artificielle ne peut pas le faire et si elle le faisait nous aurions un autre problème pour l'avenir de l'humanité, autrement plus important que ceux auxquels nous faisons face aujourd'hui. 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 entièrement 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 acceptable, socialement responsable, voire souhaitable.**



# 1. POURQUOI LA DÉCLARATION DE MONTRÉAL IA RESPONSABLE?

La *Déclaration de Montréal* pour un développement responsable de l'IA est une œuvre collective qui a pour objectif de mettre le développement de l'intelligence artificielle au service du bien-être collectif et individuel, 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 **VALEURS FONDAMENTALES**.



Le travail initial d'identification de ces valeurs et de ces principes permet de lancer un processus de participation citoyenne qui précisera 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.

## 1.1

### 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<sup>e</sup> 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 maintenant 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 prenons 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.

L'utilisation d'algorithmes pour prendre des décisions techniques ou administratives n'est pas nouvelle. 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. 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, d'une perte d'autonomie et d'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 ?

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.

<sup>1</sup> Paul Meehl, *Clinical versus Statistical Prediction*, University of Minnesota, 1954.

## 1.2

### 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 et lancer les travaux de la *Déclaration de Montréal pour un développement responsable de l'IA*<sup>2</sup>. 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<sup>3</sup> 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.

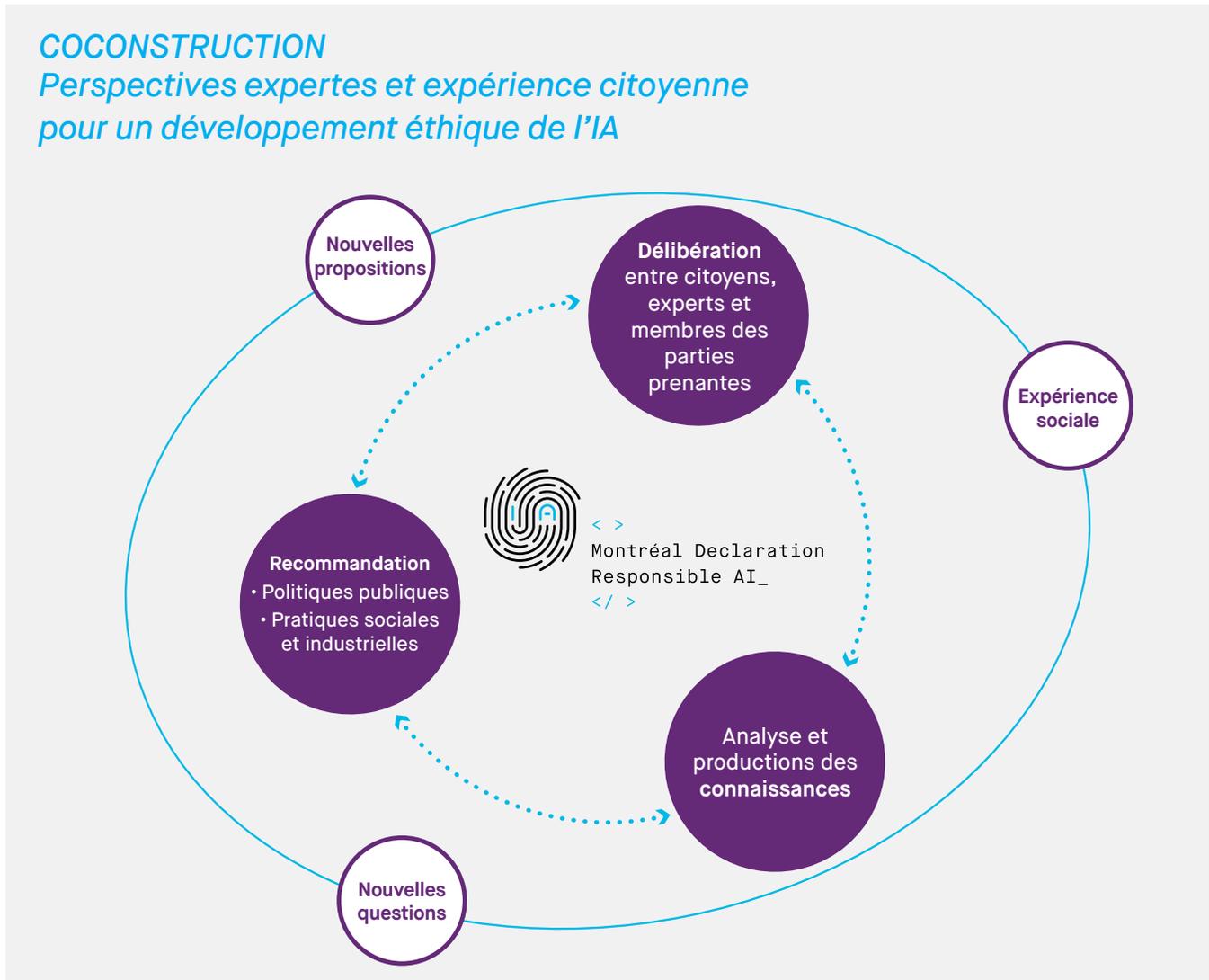
<sup>2</sup> Le comité scientifique du Forum était composé de Louise Béliveau (UdeM, Vice-rectorat aux affaires étudiantes et aux études), Yoshua Bengio (UdeM, Département d'informatique, MILA, IVADO), David Décary-Héту (UdeM, École de criminologie), Nathalie De Marcellis-Warin (École Polytechnique, Département de mathématiques et de génie industriel, CIRANO – Centre interuniversitaire de recherche en analyse des organisations), 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).

<sup>3</sup> 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).

### 1.3

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

Figure 2 : La démarche de coconstruction



Tel que mentionné en début de chapitre, le travail initial d'identification de ces valeurs et des principes correspondants a pour seul objectif de lancer un processus de participation citoyenne qui permette 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 *Déclaration de Montréal* 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.

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. 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.

Une fois le processus de coconstruction achevé, 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 et, 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 d'éthique ou de déontologie, ou qu'elles 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 de Montréal* pour un développement responsable de l'IA 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.

## 1.4

### MONTRÉAL ET LE CONTEXTE INTERNATIONAL

L'initiative de la *Déclaration de Montréal* s'inscrit dans un contexte scientifique, social et industriel favorable. Montréal est devenu un pôle de recherche et de développement majeur en intelligence artificielle avec une communauté de chercheurs (Yoshua Bengio à l'Université de Montréal, pionnier dans le domaine de l'apprentissage profond, Joëlle Pineau et Doina Precup à l'Université McGill, Hugo Larochelle et Simon Lacoste-Julien à l'Université de Montréal, et tant d'autres), des laboratoires universitaires de réputation mondiale (MILA, IVADO) et une pépinière de start-ups et d'entreprises en plein essor (Element AI, Imagia pour ne citer que deux exemples). 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 de Montréal<sup>4</sup>, est aussi ce 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 offre 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

<sup>4</sup> [http://ville.montreal.qc.ca/portal/page?\\_pageid=5798,141982209&\\_dad=portal&\\_schema=PORTAL](http://ville.montreal.qc.ca/portal/page?_pageid=5798,141982209&_dad=portal&_schema=PORTAL)

données massives) à Québec constituent des pôles d'excellence dans la recherche fondamentale qui ont entraîné une croissance industrielle extrêmement rapide et robuste. Tout comme les Fonds de recherche pour le Québec, l'Institut canadien de recherches avancées (ICRA) 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 de Montréal* est ainsi portée par différents acteurs québécois et canadiens hors Montréal.

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<sup>5</sup> et l'EGE (*European Group on Ethics in Science and New Technologies*<sup>6</sup>) 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<sup>7</sup>.

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<sup>8</sup>. 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? ». Sans prétendre à l'exhaustivité, mentionnons enfin la démarche participative de la CNIL (Commission nationale de l'informatique et des libertés) en France qui a abouti à 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 ». 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.

<sup>5</sup> Nous tenons à remercier le *UK Science and Innovation Network in Canada* qui a facilité ce dialogue.

<sup>6</sup> 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.

<sup>7</sup> Entrevue de Yoshua Bengio lors de la conférence d'Asilomar : [futureoflife.org/2017/01/18/yoshua-bengio-interview/](http://futureoflife.org/2017/01/18/yoshua-bengio-interview/)

<sup>8</sup> AI Now Institute, *Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability*, Avril 2018.

La particularité de la démarche de la *Déclaration de Montréal* est d'être essentiellement participative. De février à avril 2018, le processus de coconstruction a mobilisé plus de 500 citoyens, experts et parties prenantes au cours d'une quinzaine d'ateliers et de journées de coconstruction. 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.

<sup>9</sup> EGE, *Artificial Intelligence, Robotics and 'Autonomous' Systems* (mars 2018), p. 14.



## 2. LA VERSION PRÉLIMINAIRE DE LA DÉCLARATION DE MONTRÉAL POUR UN DÉVELOPPEMENT RESPONSABLE DE L'IA



&lt; &gt;

# Déclaration de Montréal IA responsable\_

&lt; / &gt;

## LA DÉCLARATION DE MONTRÉAL POUR UN DÉVELOPPEMENT RESPONSABLE DE L'IA (VERSION PRÉLIMINAIRE)

### PRÉAMBULE

Qu'elle soit naturelle ou artificielle, l'intelligence n'a pas de valeur en soi. L'intelligence d'un individu ne nous dit rien de sa valeur morale; c'est aussi le cas pour toute autre entité intelligente. L'intelligence peut néanmoins avoir une valeur instrumentale : c'est un outil qui peut nous éloigner ou nous rapprocher d'un objectif que nous valorisons. Ainsi, l'intelligence artificielle (IA) peut créer de nouveaux risques et exacerber les inégalités économiques et sociales. Mais elle peut aussi contribuer au bien-être, à la liberté ou à la justice.

D'un point de vue éthique, le développement de l'IA pose des défis inédits. Pour la première fois dans l'histoire, nous avons la possibilité de créer des agents non humains, autonomes et intelligents, qui n'ont pas besoin de leurs concepteurs pour accomplir des tâches que l'on croyait réservées à l'esprit humain. Ces machines intelligentes ne se contentent pas de mieux calculer que les êtres humains, elles cherchent, traitent et diffusent des informations. Elles interagissent avec des êtres sensibles, humains ou non humains. Bientôt, elles pourront même leur tenir compagnie, comme un parent ou un ami.

Ces agents artificiels vont donc être amenés à influencer directement nos vies. À long terme, on pourrait créer des « machines éthiques », c'est-à-dire capables de prendre des décisions en se conformant à des principes éthiques. Il faut se demander si ces développements sont responsables et souhaitables. Et il est permis d'espérer que l'IA rende nos sociétés meilleures, dans l'intérêt de tous et le respect de chacun.

# VALEURS ET PRINCIPES

## BIEN-ÊTRE

Le développement de l'IA devrait ultimement viser le bien-être de tous les êtres sentients.

## AUTONOMIE

Le développement de l'IA devrait favoriser l'autonomie de tous les êtres humains et contrôler, de manière responsable, celle des systèmes informatiques.

## JUSTICE

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.

## VIE PRIVÉE

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.

## CONNAISSANCE

Le développement de l'IA devrait promouvoir la pensée critique et nous prémunir contre la propagande et la manipulation.

## DÉMOCRATIE

Le développement de l'IA devrait favoriser la participation éclairée à la vie publique, la coopération et le débat démocratique.

## RESPONSABILITÉ

Les différents acteurs du développement de l'IA devraient assumer leur responsabilité en œuvrant contre les risques de ces innovations technologiques.



### 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 la *Déclaration de principes éthiques* 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 17<sup>e</sup> 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<sup>10</sup>. 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 »<sup>11</sup>, 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 apparaît 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 21<sup>e</sup> siècle pour assister au printemps de l'IA.

L'IA consiste à simuler certains processus d'apprentissage de l'intelligence humaine<sup>12</sup>, à s'en inspirer et à les 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 des informations 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 qui permettent à la machine de percevoir son environnement et d'élaborer une représentation de cet environnement.

<sup>10</sup> Leibniz (1666), *De Arte combinatoria* (« De l'art combinatoire »).

<sup>11</sup> A. M. Turing (1950), « Computing Machinery and Intelligence ». *Mind* 49, p. 433-460.

<sup>12</sup> 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. »

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 donc d'abord 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 internet et une multitude de capteurs qui produisent en tout temps des données qui sont non structurées, c'est-à-dire qui sont issus de supports d'information désordonnés qu'il faut manipuler et transformer avant de pouvoir les exploiter. 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'une grande quantité d'images.

Ce qui a aussi 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 couplés à 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 : l'algorithme apprenant produit ses propres modèles pour traiter les données et trouver la solution.. 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 éloquentes, 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<sup>13</sup>. 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 :

« Bonjour, je m'appelle Inès, 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 *chatbots* pour demander des informations sur son assurance maladie ou un nouveau plan bancaire, ou encore pour demander un conseil vestimentaire.

Pour l'instant, les *chatbots* 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 ? »<sup>14</sup>. Or, dès l'introduction de son article, il change le problème auquel il estime pouvoir donner une

<sup>13</sup> 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).

<sup>14</sup> A. M. Turing (1950).

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 ».

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<sup>15</sup>. 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 *chatbots* 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 et 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, à 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. 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 en soi bonne ou mauvaise, 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.

<sup>15</sup> J. Searle (1980), 'Minds, Brains and Programs'. *Behavioral and Brain Sciences* 3, p. 417-57.

### 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 réglementation, 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<sup>16</sup>? 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 *chatbots* 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.

<sup>16</sup> 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 améliorer le système de santé pour que tout le monde puisse consulter un médecin ? Il n'y a aucune réponse simple, mais il y a des choix à faire.

La *Déclaration de Montréal* est actuellement dans une version préliminaire qui sert de déclencheur à la réflexion éthique. Les valeurs qui ont été présentées dans cette version, quoiqu'elle soit incomplète, fournissent un vocabulaire moral de base pour amorcer l'analyse éthique de situations banales et pour faciliter la délibération. 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, la justice. Il serait aisé de montrer que la valeur de vie privée permet de situer le problème de la confidentialité des données du patient.

Les notions précédentes ont été présentées dans les différentes journées et ateliers de coconstruction<sup>17</sup>. Elles ont servi de points de départ à la délibération sur des scénarios prospectifs et à l'élaboration de

pistes de solution éthiques, politiques et juridiques aux enjeux sociétaux de l'IA. Ces exposés informatifs se terminaient sur la question générale suivante :

***Comment doit-on organiser la société pour faire un usage éthique des machines intelligentes, compatible avec nos intérêts moraux et sociaux fondamentaux ? Quelles règles devrait-on suivre pour faire le meilleur usage de ces machines tout en protégeant notre autonomie, en garantissant l'égalité sociale et en répartissant équitablement les fruits de l'économie de l'IA ?***

Le processus de coconstruction a permis d'apporter différentes réponses crédibles à cette interrogation.

<sup>17</sup> D'autres analyses ont été présentées, notamment à partir du cas des véhicules autonomes. Nous les exposerons de manière plus complète dans la version finale de ce rapport.



# 4. LA DÉMARCHÉ DE COCONSTRUCTION

## 4.1

### LES PRINCIPES DE LA DÉMARCHÉ 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 documentées et crédibles aux questions pressantes et la formulation de recommandations politiques et juridiques disposant d'une forte légitimité démocratique.

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 (coconstruction) 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 :

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.

## 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.

## 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.

## 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.

## 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.

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 proposons de le faire. 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.

Pour toutes ces raisons, nous avons retenu 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.

## 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.

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 lesquels ils sont erronés.

Le rôle des éthiciens est donc celui d'un accompagnement éclairé. 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* IA Responsable 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.



Plus de 10 ateliers de coconstruction ont été organisés de février à mai : 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, puis au Musée de la civilisation à Québec).

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<sup>18</sup>. En passant du modèle de l'espace de prêt documentaire à celui de la « bibliothèque tiers-lieux » inclusive et cherchant à renforcer les capacités 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.

<sup>18</sup> 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

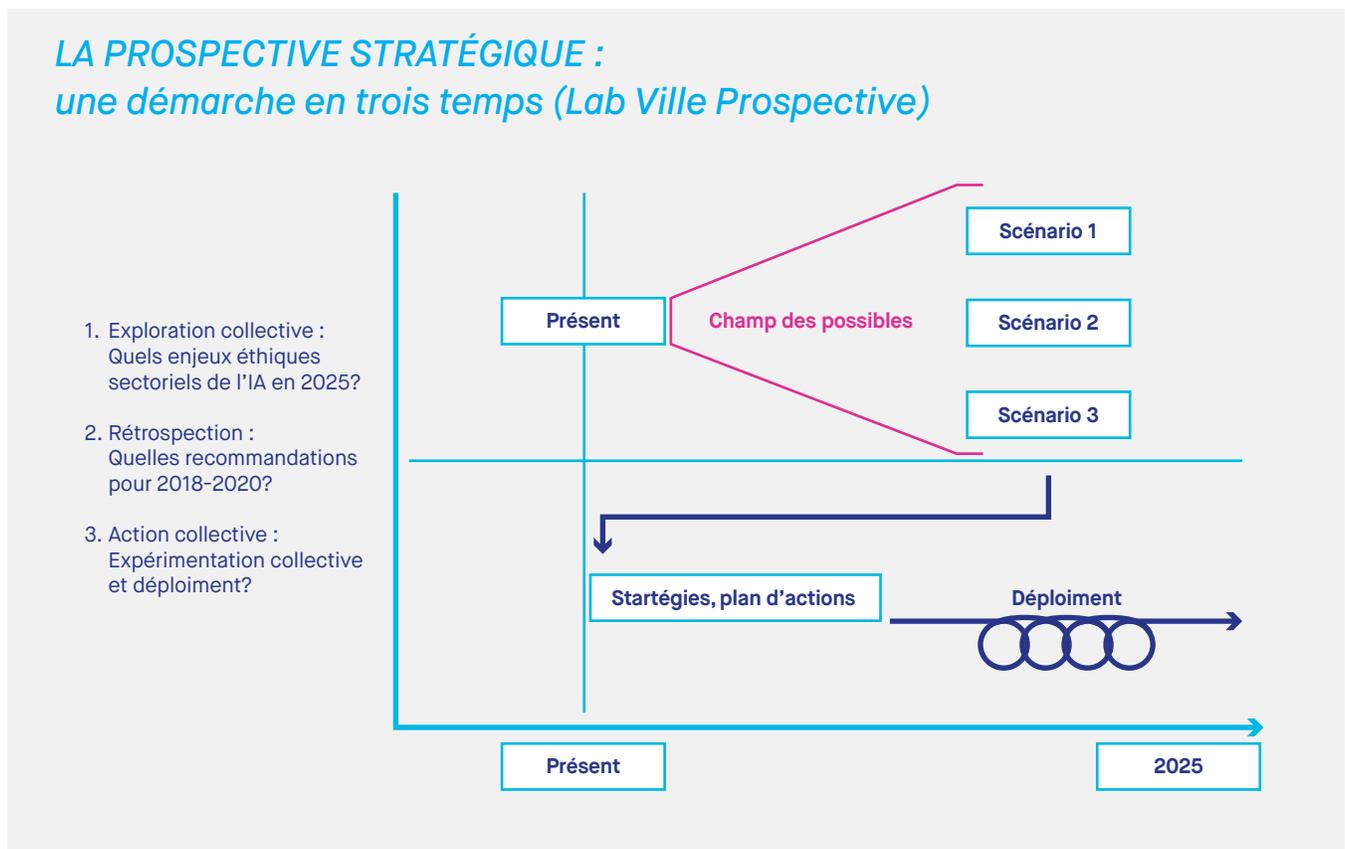
### ORIGINALITÉ DE LA DÉMARCHÉ DE COCONSTRUCTION

Au regard des autres initiatives en éthiques 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 3 : La prospective stratégique : une démarche en trois temps



- > Ensuite, le recours à des méthodes d'animation de design participatif en « forum hybride »<sup>19</sup> 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 (ex. site 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 sur les 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.

<sup>19</sup> Callon, Lacoumes, Barthe, *Agir dans un monde incertain. Essai sur la démocratie technique*, Paris, Le Seuil, 2001

## 4.5

### PORTRAIT DES PARTICIPANTS

Le recrutement de citoyens, d'experts et de professionnels de différents secteurs du marché du travail a permis d'atteindre 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.

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 4 : proportion hommes-femmes ayant participé aux ateliers de coconstruction

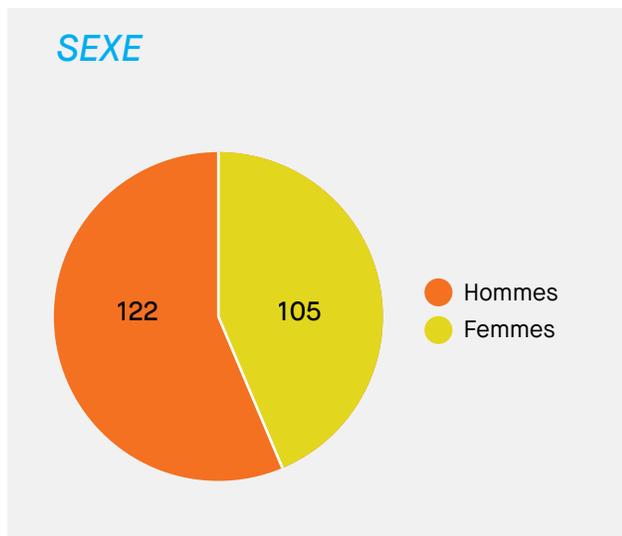


Figure 5 : Les participants aux ateliers de coconstruction par tranches d'âge

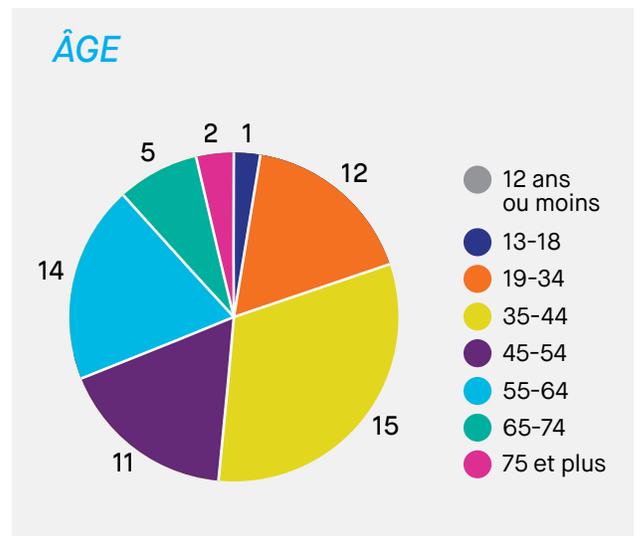


Figure 6 : Répartition des répondants aux cafés citoyens et aux journées de coconstruction par niveau de scolarité atteint

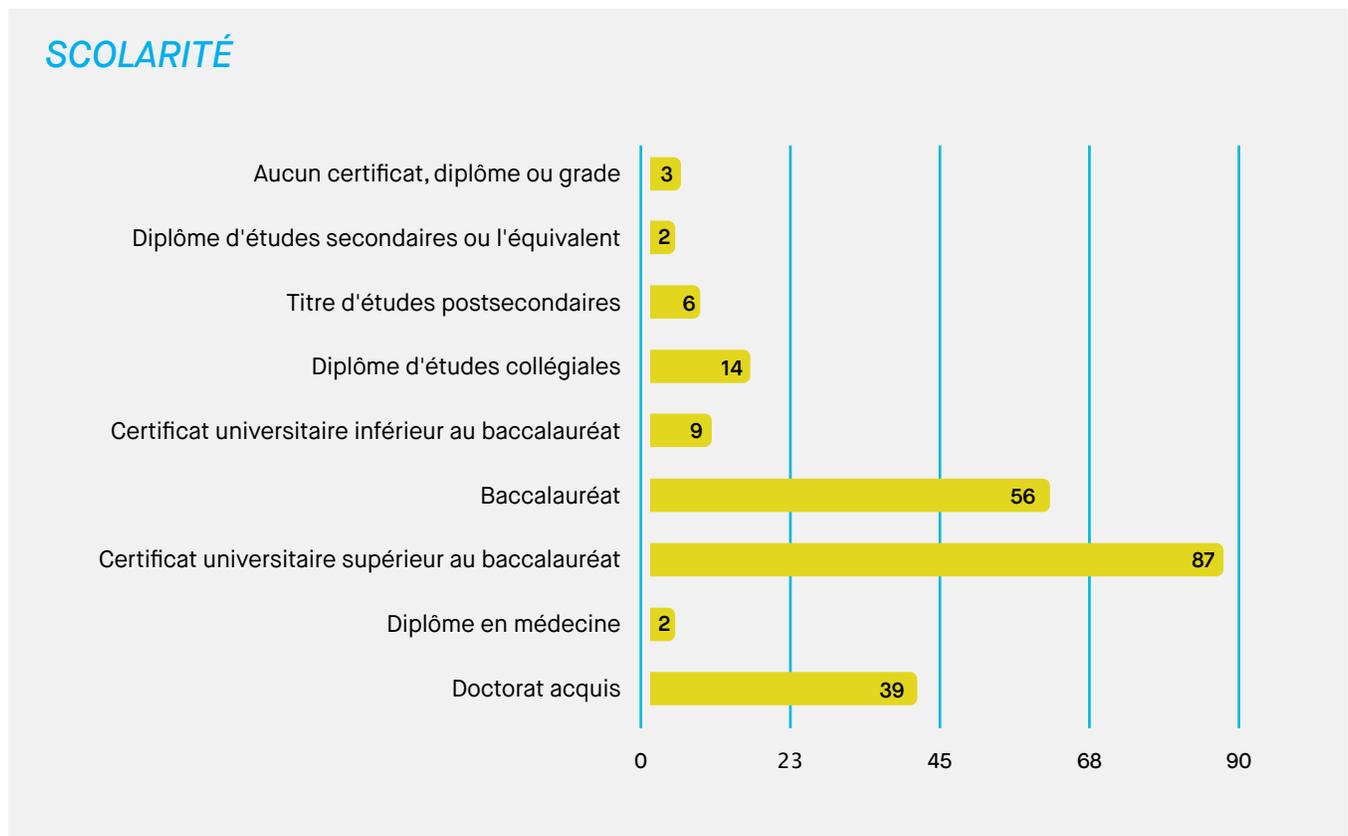
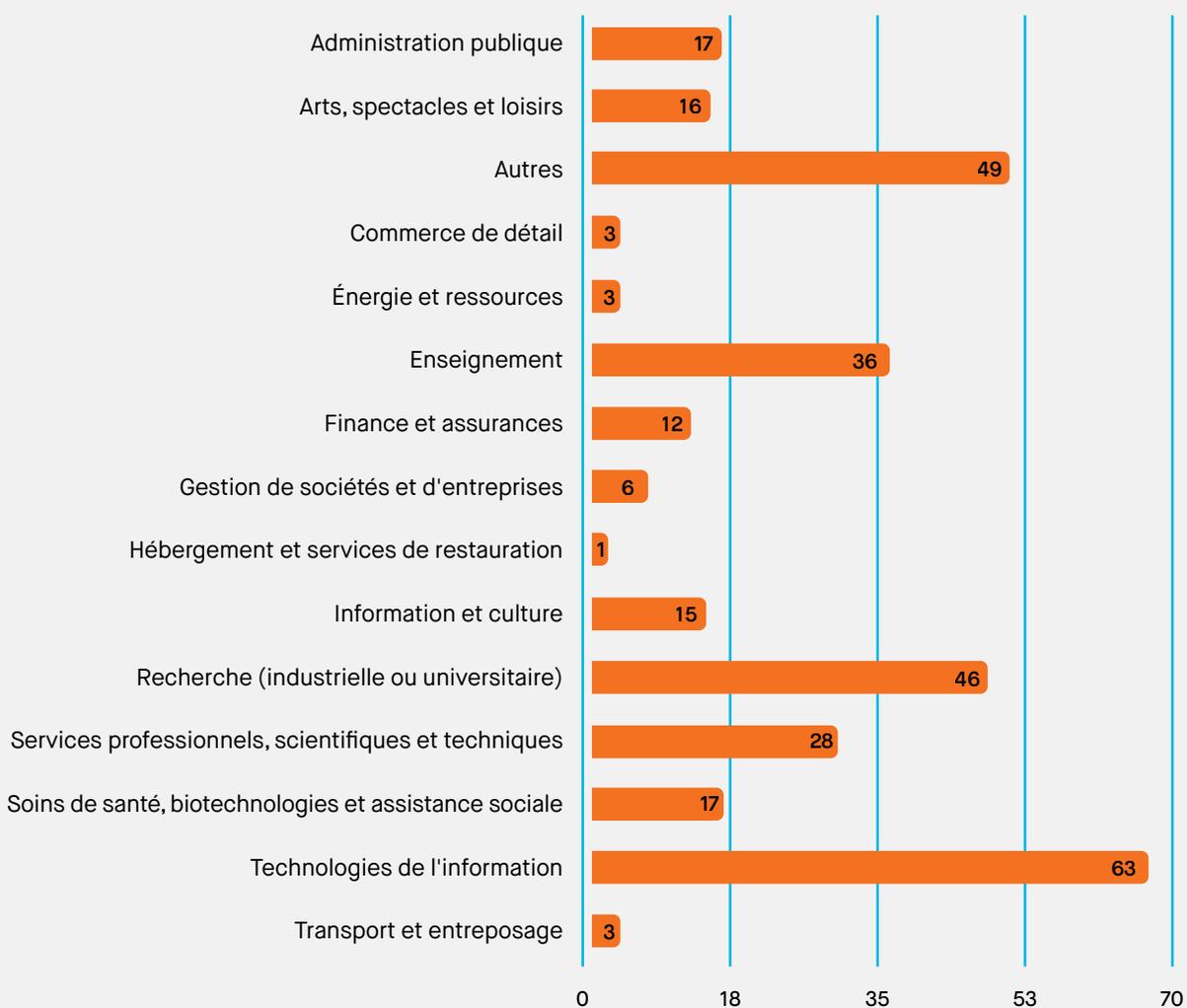


Figure 7: Répartition des répondants aux cafés citoyens et aux journées de coconstruction par secteur d'activité

## SECTEURS D'ACTIVITÉ

34% des répondants ont indiqués 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-its 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*<sup>20</sup> (Politix, 2002, p. 25-26) : rendre possible l'imagination de solutions nouvelles dans un monde incertain; permettre une montée 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.

<sup>20</sup> 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 et citoyennes 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 » de la Gazette 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-décision 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 ? », 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
<b>Description</b>	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.
<b>Principes associés</b>	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 » de la Gazette 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<sup>21</sup>.

<sup>21</sup> 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 » de la Gazette 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 *co-bot* 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 responsabilité 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
<b>Description</b>	<p>Trop de concentration de pouvoir (cf. GAFAM) ne permettant pas :</p> <ul style="list-style-type: none"> <li>- Le partage équitable des bénéfices de l'IA</li> <li>- L'entrée de nouveaux joueurs (nouveaux modèles d'affaire de type COOP)</li> <li>- diminuer les inégalités (littératie)</li> </ul>	<p>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.</p>	<p>Définir le bien commun et le type de responsabilité collective dans la transition numérique</p> <p>Par exemple : quelles parties prenantes? L'entreprise seule? L'État? Les syndicats? Le système éducatif?</p>
<b>Principes associés</b>	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
<b>Description</b>	<p><b>Renforcer la littératie numérique pour tous.</b></p> <p>Avec un référentiel de compétences, les bibliothèques publiques, les écoles, en milieu de travail. En traitant la question de l'analphabétisme et du «non-recours» des citoyens.</p>	<p><b>Comité permanent interministériel mixte sur l'IA, exécutif à côté du PM.</b></p> <p>À l'interface des thèmes de l'économie, de l'emploi, de l'éducation et de la culture (cf. Stratégie numérique)</p>	<p><b>Fonds d'assurance numérique sur l'IA pour permettre de se former et s'adapter.</b></p> <p>Exemple de dispositif: le Régime d'assurance parentale 50 semaines, qui peut aussi inspirer un revenu minimum contre la précarisation</p>	<p><b>Incitatifs sur les nouveaux modèles d'entreprises de traitement des Data</b></p> <p>Exemple : COOP pour rompre l'isolement de travailleurs autonomes opérant sur les data et assurer une autonomie collective</p>	<p><b>Orientation des investissements vers l'IA responsable pour le bien commun</b></p> <p>Modèle de l'ISR (Investissement socialement responsable). Investissement de l'État, des particuliers, en synergie avec les Fonds de Travailleurs</p>	<p><b>Processus accéléré de mise à jour et de création de programmes professionnels</b></p> <p>Avec cégeps, universités, ministères, ordres professionnels impactés par l'IA (ex. droit, santé)</p>
<b>Catégories d'instrument</b>	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 » de la Gazette 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* l'IA responsable, 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 renouvellements 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.



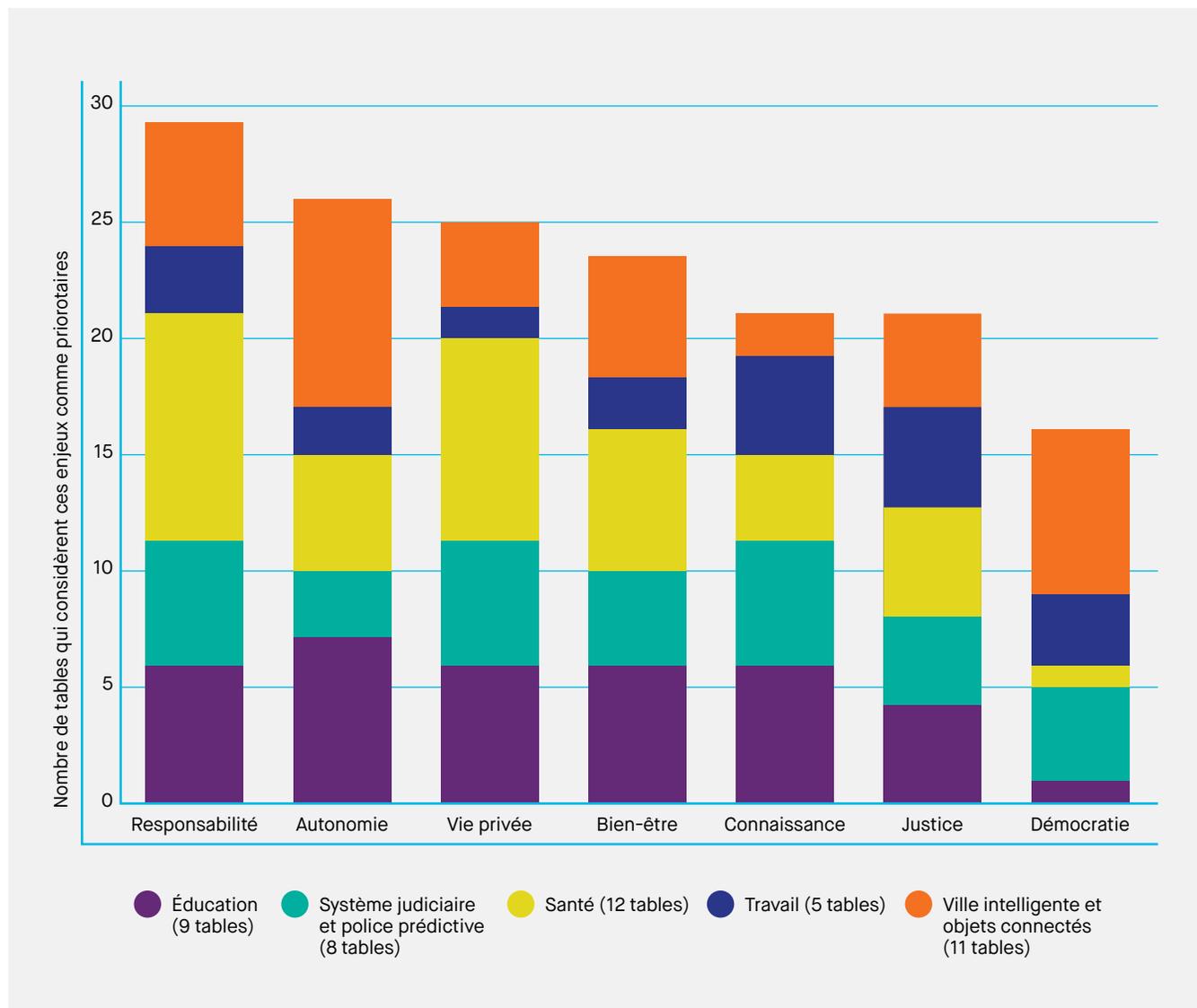
# 6. LES PREMIERS RÉSULTATS DE LA COCONSTRUCTION

## 6.1

### RÉSUMÉ

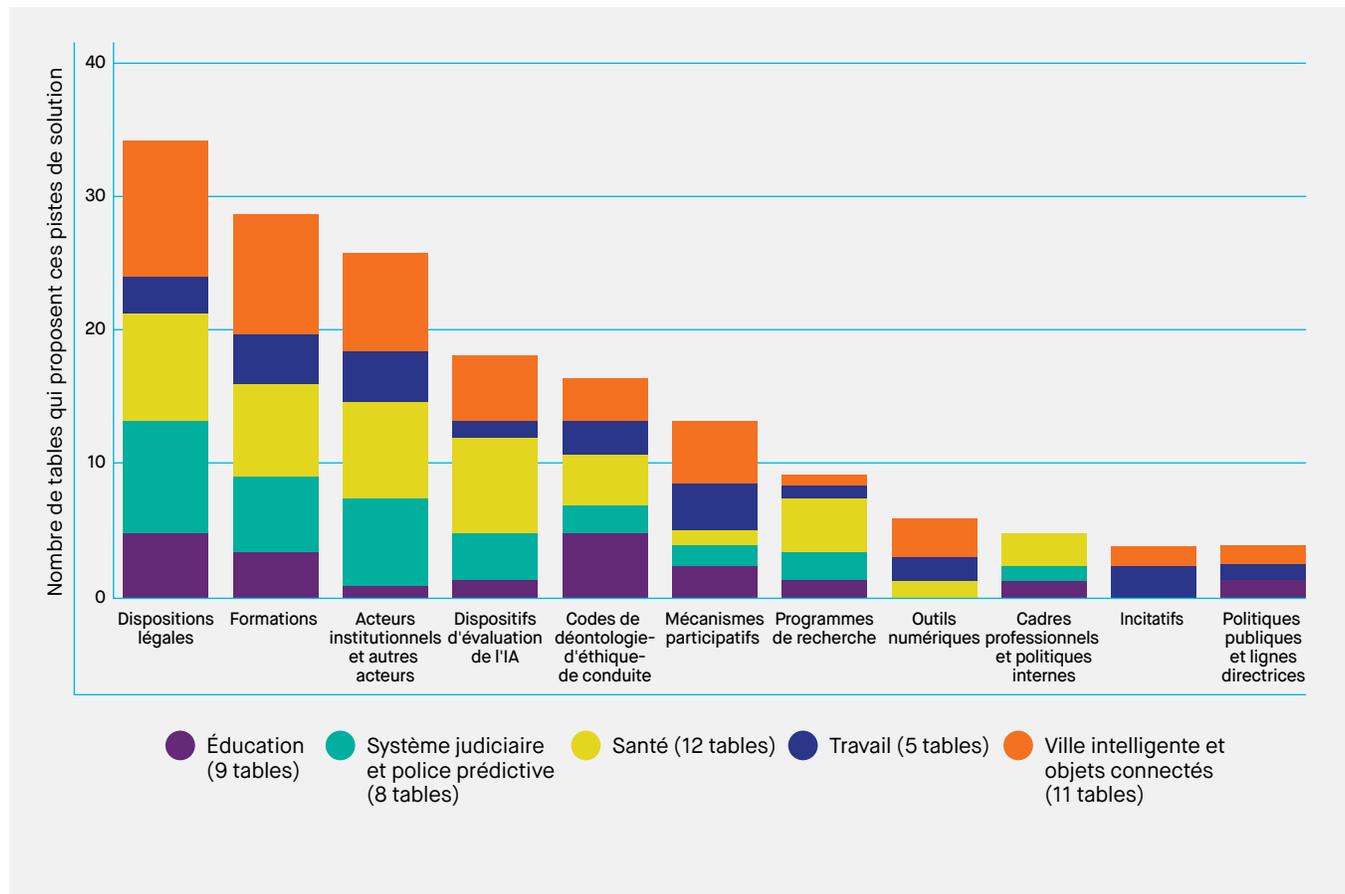
Les citoyens se sont réunis autour de 45 tables pour discuter de leur perception des enjeux liés à l'application des principes de la Déclaration.

Tableau 5 : les principes auxquels réfèrent les enjeux prioritaires identifiés par les citoyens.



Ils ont identifié différentes catégories de pistes de solution pour répondre à ces enjeux.

Tableau 6 : les pistes de solution proposées pour répondre aux enjeux identifiés



## 6.2

### 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 des 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).

Ces résultats sont issus d'une analyse préliminaire et non exhaustive des deux principaux axes discutés autour de chaque table : les différents enjeux que soulève le développement de l'IA ainsi que des pistes de solution identifiées pour y répondre qui sont ressorties des discussions citoyennes sur la base de scénarios déclencheurs. À ce stade, cette analyse est restée descriptive et au plus proche de la parole des citoyens. Pour les fins de ce rapport, l'accent a été mis sur : 1) les grandes directions attendues en termes de développement responsable de l'IA; 2) la présentation des enjeux que les citoyens ont considérés comme prioritaires; 3) les enjeux qui pourraient mener à la création de nouveaux principes dans la Déclaration; 4) les pistes de solutions identifiées par les citoyens pour répondre à ces enjeux.

Les grandes directions attendues en termes de développement responsable de l'IA réfèrent aux recommandations des citoyens qui ne se sont pas précisées en pistes de solution concrètes. Elles permettent néanmoins de dégager les principales positions et attentes normatives des citoyens face au développement de l'IA.

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 analysés pour les fins de ce rapport. Ces enjeux prioritaires ont donc été décrits sur la base des formulations citoyennes et classés, pour chaque secteur, en fonction des principes 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 est détaillé dans ce rapport d'étape.

Différents enjeux qui pourraient mener à la création de nouveaux principes dans la Déclaration ont été identifiés sur la base de l'ensemble des discussions qui ont eu lieu. Sont présentés dans ce rapport, de manière non exhaustive, ceux qui se sont avérés particulièrement pertinents.

Enfin, les pistes de solution identifiées par les citoyens pour répondre à ces enjeux ont été classées en 11 principales catégories. Ces catégories seront amenées à être précisées lors des étapes subséquentes de l'analyse. La catégorie qui semblait la plus pertinente à aborder pour chacun des secteurs est présentée plus en détail.

Concernant les données quantitatives de ce rapport, le nombre d'occurrences correspond au nombre de tables où chaque enjeu/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, des pistes de solution émises lors des discussions mais n'apparaissant pas comme telles sur les affiches, ont également été prises en considération.

Les citations du rapport sont présentées de façon à faire référence à la table de coconstruction lorsqu'elles sont issues de formulation 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).

## 6.3

### DÉVELOPPEMENT RESPONSABLE DE L'IA : 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. »

- Un membre participant

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. »

- Un participant

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 legalistes 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 mentionné 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. » - Un participant.

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. » - Un participant.

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.

## 6.4

### RÉSULTATS : LA PERCEPTION CITOYENNE DES ENJEUX

Les grands enjeux du développement responsable de l'IA

Tableau 7 : 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
<b>Responsabilité</b>	6	5	3	10	5	29
<b>Autonomie</b>	7	3	2	5	9	26
<b>Vie privée</b>	6	5	1	9	4	25
<b>Bien-être</b>	6	4	2	6	5	23
<b>Connaissance</b>	6	5	4	4	2	21
<b>Justice</b>	6	4	5	4	4	21
<b>Démocratie</b>	1	4	3	1	7	16
<b>Nombre total de tables de coconstruction</b>	9	8	5	12	11	45

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'intelligence artificielle.

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.

## 6.4.1

### 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 : 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). Le manque de transparence des algorithmes vient alors, 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).

## 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 la qualité 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 relatifs à l'atteinte à la vie privée et à la confidentialité. Ces enjeux concernent une potentielle invasion dans la vie privée, qui peut être liée au développement et à la configuration des systèmes d'IA (ex. 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 adressés 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 à 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).

## 6.4.2

### DES ENJEUX QUI POURRAIENT CONDUIRE À LA CRÉATION DE NOUVEAUX PRINCIPES

Différents enjeux identifiés ou discutés par les citoyens semblent mériter un intérêt particulier et pouvoir mener à la création d'éventuels nouveaux principes dans la *Déclaration de Montréal*, notamment pour leur aspect transversal (tant au niveau des secteurs que des principes).

#### DES ENJEUX ENVIRONNEMENTAUX

Par exemple, l'impact du développement et de l'utilisation responsable de l'IA sur l'environnement. Ces enjeux interrogent la façon de garantir l'utilisation responsable et équitable des ressources matérielles et naturelles. Ils posent également la question d'assurer un bilan énergétique positif en ce qui concerne les effets polluants de l'IA et des technologies associées à son fonctionnement.

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

- Un participant

#### PRÉCISER LE PRINCIPE DE JUSTICE : DIVERSITÉ ET ÉQUITÉ

Le principe de justice a été abordé selon deux types d'enjeux : tantôt en termes de diversité, tantôt en termes d'équité et de justice sociale. Ce principe pourrait donc être scindé en deux pour mettre de l'avant chacun de ces aspects :

- > Un principe de diversité pourrait ainsi viser à éviter les discriminations en trouvant des mécanismes dépourvus de biais liés au genre, à l'âge, aux capacités mentales et physiques,

à l'orientation sexuelle, aux origines ethniques et sociales et aux croyances religieuses, sans en créer de nouveaux. Le principe de diversité appelle également à favoriser la multitude des perspectives et des intelligences plutôt que la normalisation de profils d'individus selon un nombre restreint de catégories et de critères.

**« Une perte de la diversité entraîne une standardisation extrême. On revient au besoin de conserver la complexité pour traiter des questions humaines. »**

- Un participant

- > Un principe d'équité ou de justice sociale impliquerait que les bénéfices de l'IA soient accessibles à toutes et tous, et que le développement de l'IA ne contribue pas à l'accroissement des inégalités économiques et sociales, mais qu'il les réduise.

**« L'accessibilité aux nouvelles technologies : privilège des plus riches ? Est-ce que ce type de technologie creusera les inégalités ? »**

- Un participant

#### DES ENJEUX DE CONFIANCE, FIABILITÉ, SÉCURITÉ : UN PRINCIPE DE PRUDENCE?

Par ailleurs, les enjeux relatifs à la confiance envers le développement des technologies d'IA ont régulièrement été soulevés. L'enjeu de la confiance en l'IA et ses algorithmes dans différents secteurs se pose principalement sous l'angle d'une certaine méfiance envers ces techniques ainsi qu'à l'égard de la représentativité des données sélectionnées et la validité des interprétations qui en sont faites, invitant à un principe de prudence :

**« Comme c'est scientifique, l'individu pourrait tendre à négliger que l'algorithme peut se tromper : prudence. »** - Un participant

Cet enjeu de confiance entretient ainsi une relation étroite avec la question de la fiabilité des systèmes d'IA. Il invite à porter attention à la qualité des données collectées et des corrélations qui en sont faites ainsi qu'à leur visée pour ne pas s'en tenir à une confiance aveugle et pour prévenir de potentiels effets de manipulation.

Dans ce même ordre d'idées, les participants ont soulevé des enjeux liés à la sécurité des dispositifs d'IA, notamment face au risque de potentiels abus, piratages et cyberattaques des systèmes et des données qu'ils contiennent, ou à la validité des recommandations et décisions issues des algorithmes. Ces enjeux s'accompagnent d'un dilemme entre la « fluidité collective » et la « vulnérabilité du système », soit une IA qui doit être à la fois malléable et robuste (ex. dans le secteur ville intelligente et objets connectés).

## VERS UN PRINCIPE DE TRANSPARENCE?

À la croisée des principes de connaissance, de responsabilité et de justice se dessine un principe de transparence qui 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 :

**« 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).

Cet enjeu d'explicabilité implique la nécessité de trouver une forme de vulgarisation de ces procédures algorithmiques pour les rendre intelligibles à chacun ; il va de pair également avec le développement d'une littératie numérique qui permette d'avoir un esprit critique sur ce système et d'agir par consentement éclairé. L'explicabilité de ces algorithmes accompagne l'enjeu de pouvoir vérifier les décisions algorithmiques, attribuer la responsabilité de ces décisions et éventuellement corriger certains effets néfastes comme des biais discriminatoires. Il s'agit par ailleurs de rendre accessibles ces explications d'algorithmes, par exemple par le développement en mode ouvert (open source, licence libre, données ouvertes), notamment dans un souci de rétroaction pour comprendre pourquoi telle décision a été prise et gérer d'éventuels sentiments d'injustice face à des refus (dans le cadre d'accès à un emploi, à une aide sociale ou une assurance, par exemple).

Une mise en garde a cependant été faite à deux reprises quant à ce principe de transparence : cette transparence pourrait avoir un potentiel effet sur la sécurité des systèmes algorithmiques (risque de hacking). Ce principe de transparence est ainsi voisin des enjeux de confiance accordée aux technologies d'IA.

**« Si l'IA analyse des choses trop complexes pour l'esprit humain, qui vérifie ce qui se passe derrière le rideau ? »**

**- Un participant.**

### 6.4.3

## DES ENJEUX RELATIFS AU RAPPORT ENTRE HUMAINS ET IA

Quel que soit le secteur, les citoyens ont identifié de nombreux enjeux relatifs au rapport entre êtres humains et IA. Les participants s'inquiètent notamment de la place que pourra occuper l'humain avec l'avènement de tels systèmes, ce qui pourrait conduire à différentes dérives.

Par exemple, les citoyens se préoccupent du respect de la « nature humaine », et ce quels que soient les secteurs. Celle-ci vient se confronter à la place de l'objet dans la société et son rapport à l'humain pour le secteur ville intelligente et objets connectés : va-t-on diminuer le statut de l'humain et accorder plus d'importance à la protection des objets ? Respecter la « nature humaine » signifie également pour les citoyens de garantir la prise en compte de la singularité, de la complexité, voire de la désorganisation humaine selon de nombreux paramètres difficilement quantifiables, par exemple ce que les participants ont appelé « le dynamisme individuel » dans le secteur judiciaire. 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. En santé, cette fiabilité est également remise en question en ce qui concerne le diagnostic ou les conseils issus d'algorithmes qui n'ont pas de vision holistique des individus alors que ceux-ci ne peuvent être réduits à leurs données biologiques.

Les participants s'inquiètent ainsi d'une éventuelle déshumanisation des services qui pourrait apparaître si une place trop importante est accordée à l'IA. En santé, il s'agit d'une certaine déshumanisation du soin et d'une perte du lien dans la relation médecin-patient. Dans le secteur ville intelligente et objets connectés, les participants s'inquiètent de l'équilibre à trouver pour assurer un développement harmonieux de la société et des êtres vivants tout en implantant de l'IA et des objets connectés. Dans le secteur du monde du travail, cette déshumanisation peut être perçue comme une mécanisation des

tâches. Dans le secteur judiciaire, elle pourrait venir de l'absence potentielle « d'empathie », « d'instinct », « de sagesse » des systèmes d'IA, ce qui amène des inquiétudes relatives à une **judiciarisation** des cas plutôt que de les traiter « humainement » :

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

- Un participant

Les participants se préoccupent ainsi d'une perte de la qualité relationnelle et émotionnelle, parfois perçue comme une potentielle « dénaturalisation », voire une « aliénation » (de la vie sociale au profit de la vie numérique), et ce dans tous les secteurs. Ces préoccupations réfèrent notamment à la transformation du rapport au soin, à la connaissance, au savoir, au travail, mais aussi aux compétences des individus.

« Le défi n'est pas de rendre les machines plus intelligentes, mais de rendre plus intelligents les humains. »

- Un participant

Les médecins auront-ils toujours la même expertise s'ils s'en remettent constamment à des systèmes experts ? Quel effet cela aurait-il sur la confiance qu'ils accorderont à leur expertise face à celle de l'IA ? Une réflexion similaire a été soulevée dans le secteur de l'éducation par rapport au remplacement des enseignants par les IA :

« Dans les écoles, s'il y a trop d'équipements d'IA, les professeurs vont devenir inutiles. »

- Un participant

Toujours dans le secteur de l'éducation, les citoyens ont ainsi rappelé qu'il ne faut pas oublier que l'intervention humaine est nécessaire :

« On ne peut pas se fier uniquement à la machine. »

- Un membre participant

Trouver la voie d'une complémentarité entre humain et IA semble ainsi d'une grande importance. Cette complémentarité a été abordée comme « un bon équilibre » pour le partage des tâches (par exemple, entre l'« objectif » et le « subjectif » pour le secteur monde du travail, entre une IA infiniment patiente comme un « assistant savoir » et un enseignant doté, lui, de capacités émotionnelles et relationnelles pour le secteur éducation). Dans le secteur du travail, les citoyens ont proposé la mise en place d'une veille pour préserver la « primauté de l'humain », qui devrait garantir que la technologie n'est qu'un soutien :

« Garantir que ce système ne soit pas une finalité en soi, mais qu'il soit centré sur l'humain. »

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

## 6.5

### RÉSULTATS : LES PISTES DE SOLUTION

#### 6.5.1

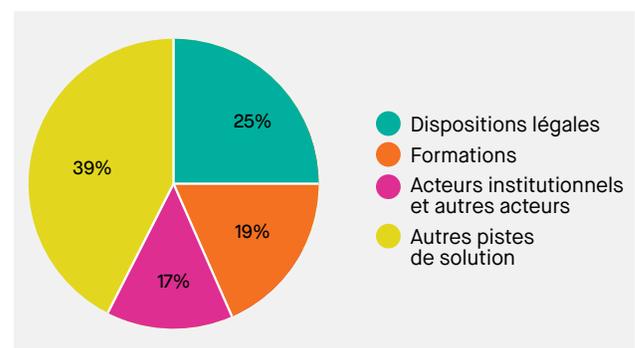
### LES GRANDES PISTES DE SOLUTIONS PROPOSÉES PAR LES CITOYENS

Les citoyens ayant participé aux journées de coconstruction ont été invités à proposer des pistes de solutions en réponse aux enjeux précédemment identifiés. 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).

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

Figure 8: Trois grandes pistes de solutions prioritaires à toutes les tables



Quel que soit le secteur, l'ensemble des tables s'est entendu pour recommander la mise en place d'un cadre légal adapté à la réalité des enjeux liés au développement de l'IA et à la gestion des données personnelles (en particulier, les données massives). Par exemple, les participants ont proposé la mise en place de lois et règlements spécifiques, de nouveaux types de contrats, voire de la mise en place d'un moratoire. 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 des objets connectés ou d'un ministère de l'éthique des données et de la protection numérique).

Pour tous les secteurs également, 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 soulevée 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 des mécanismes participatifs (ex. coconstructions, consultations publiques ou états généraux sur l'IA) afin de garantir un développement démocratique de l'IA et de sa gestion. L'importance de la mise en place de programmes de recherche dans des disciplines variées (ex. philosophie, sciences sociales, bioéthique) a également été soulevée. 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) a également été proposée.

Le développement d'incitatifs – visant 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 de subventionner les entreprises qui mettent en place des transitions pour les employés qui voient leur travail se faire remplacer par l'IA. Enfin, la mise en place de cadres professionnels (et différentes procédures internes aux entreprises) et la création de politiques publiques qui amènerait, par exemple, à la création d'une citoyenneté numérique, ont été mises de l'avant.

Tableau 8 : Nombre de tables qui proposent chacune des catégories de pistes de solutions

	Éducation	Système judiciaire et police prédictive	Santé	Monde du travail	Ville intelligente et objets connectés	Total
Dispositions légales	6	7	8	3	10	34
Formations	4	5	6	4	10	29
Acteurs institutionnels et autres acteurs	1	6	7	3	9	26
Dispositifs d'évaluation de l'IA	1	3	8	1	5	18
Codes de déontologie-d'éthique-de conduite	5	2	4	2	2	15
Mécanismes participatifs	2	2	1	3	5	13
Programmes de recherche	1	2	4	1	1	9
Outils numériques	0	0	1	2	3	6
Cadres professionnels et politiques internes	1	1	2	0	0	4
Incitatifs	0	0	0	2	1	3
Politiques publiques et lignes directrices	1	0	0	1	1	3
<b>Nombre de tables de coconstruction</b>	<b>9</b>	<b>8</b>	<b>12</b>	<b>5</b>	<b>11</b>	<b>45</b>

## 6.5.2

### PISTES DE SOLUTION PAR SECTEUR

#### ÉDUCATION

Les citoyens se sont réunis autour de 9 tables de coconstruction autour desquelles le thème du développement de l'IA dans le secteur de l'éducation a été discuté. Les participants ont formulé 27 pistes de solution ou grandes directions en termes d'encadrement de l'IA lors de ces activités.

*Tableau 9 : Pistes de solution ou grandes directions pour le secteur de l'éducation*

	Nombre de pistes de solutions formulées
Dispositions légales	8
<b>Formation</b>	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
Politiques et lignes directrices	1
Total	27

SEPT PISTES DE SOLUTION QUI CONCERNENT LA FORMATION ONT ÉTÉ FORMULÉES PAR 4 TABLES SUR 9 :

### > FORMATION

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.Sc. 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 orchestre 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 objectifs 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 accéléreraient 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).

## SYSTÈME JUDICIAIRE ET POLICE PRÉDICTIVE

Les citoyens se sont réunis autour de 8 tables de coconstruction autour desquelles le thème du développement de l'IA dans le secteur judiciaire a été discuté. Les participants ont formulé 36 pistes de solution ou grandes directions en termes d'encadrement de l'IA lors de ces activités.

*Tableau 10 : Pistes de solution ou grandes directions pour le secteur de du système judiciaire et de la police prédictive*

	Nombre de pistes de solutions formulées
<b>Dispositions légales</b>	<b>10</b>
Acteurs institutionnels et autres acteurs	7
Dispositifs d'évaluation de l'IA	5
Formation	5
Codes de déontologie-d'éthique-de conduite	2
Mécanismes participatifs	2
Programmes de recherche	2
Cadres professionnels et politiques internes	1
<b>Total</b>	<b>34</b>

DIX PISTES DE SOLUTION FORMULÉES SONT DES DISPOSITIONS LÉGALES ET SONT RECOMMANDÉES PAR 7 TABLES SUR 8 :

### > DISPOSITIONS LÉGALES

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, mais également de rendre explicites 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 que ceux-ci soient expliqués de façon intelligible. Comme premier mécanisme de transparence, plusieurs tables de participants 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 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. 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, il est par ailleurs fait mention d'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 de renforcer les 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, 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, 3 mars 2018, scénario Arrestation prédictive).

## SANTÉ

Les citoyens se sont réunis autour de 12 tables de coconstruction autour desquelles le thème du développement de l'IA dans le secteur de la santé a été discuté. Les participants ont formulé 46 pistes de solution ou attentes en termes d'encadrement de l'IA lors de ces activités.

Tableau 11 : Pistes de solution ou grandes directions pour le secteur de la santé

	Nombre de pistes de solutions formulées
Dispositions légales	11
Acteurs institutionnels et autres acteurs	9
<b>Dispositifs d'évaluation de l'IA</b>	<b>8</b>
Formation	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

HUIT PISTES DE SOLUTION FORMULÉES SONT DES DISPOSITIFS D'ÉVALUATION DE L'IA (EN SANTÉ, DES CERTIFICATIONS), ET SONT RECOMMANDÉES PAR 8 TABLES SUR 12 :

### > 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 certification (ou label) 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 Robots d'assistance aux personnes âgées.)

## MONDE DU TRAVAIL

Les citoyens se sont réunis autour de 5 tables pour formuler 32 pistes de solution en ce qui concerne le développement de l'IA dans le secteur du travail.

Tableau 12 : Pistes de solution ou grandes directions pour le secteur du monde du travail

	Nombre de pistes de solutions formulées
Dispositions légales	8
Acteurs institutionnels et autres acteurs	5
Dispositions légales	5
Incitatifs	3
<b>Mécanismes participatifs</b>	<b>3</b>
Codes de déontologie-d'éthique-de conduite	2
Outils numériques	2
Politiques et lignes directrices	2
Dispositifs d'évaluation de l'IA	1
Programmes de recherche	1
<b>Total</b>	<b>32</b>

LES PROPOSITIONS CONCERNANT LES MÉCANISMES PARTICIPATIFS SE SONT PARTICULIÈREMENT DISTINGUÉES. ELLES SONT RECOMMANDÉES PAR 3 TABLES SUR 5 :

### > 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). L'information récoltée par le numérique pourrait alors être plus accessible et cet espace serait responsable de la structuration des secteurs émergeant dans le domaine de l'emploi.

Les citoyens ont également soulevé l'importance de la participation des usagers à la conception de l'interface des outils d'IA, qui pourrait prendre la forme d'une expérimentation collective (« design

thinking ») avec les différents partenaires et qui leur permettrait de réviser le travail des programmeurs :

**« 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 devrait ainsi permettre de suivre la collecte des données et le développement des algorithmes, et de réduire les « écarts » qui pourraient porter préjudice aux individus vers les autorités compétentes (ex. comités d'éthique, corporations) pour adapter le système.

## VILLE INTELLIGENTE ET OBJETS CONNECTÉS

Les citoyens se sont réunis autour de 11 tables sur le thème du développement de l'IA dans le secteur de la ville intelligente et des objets connectés. Ces 11 tables ont formulé 51 pistes de solution.

Tableau 13 : 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	13
<b>Acteurs institutionnels et autres acteurs</b>	<b>10</b>
Formation	10
Dispositifs d'évaluation de l'IA	5
Mécanismes participatifs	5
Outils numériques	3
Codes de déontologie-d'éthique-de conduite	2
Incitatifs	1
Politiques et lignes directrices	1
Programmes de recherche	1
<b>Total</b>	<b>51</b>

### > 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 création 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, 2 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'utilisateur. 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 (Table du Musée de la civilisation, Québec, 6 avril 2018, scénario Réfrigérateur connecté).

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 (Table de l'INM, Montréal, 18 février 2018, scénario Réfrigérateur connecté).

Pour 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 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 Voiture autonome).



# 7. UN PROCESSUS DE COCONSTRUCTION EN CONTINU

## 7.1

### POURSUIVRE LA DÉLIBÉRATION

Les travaux de la *Déclaration de Montréal* se sont concentrés sur 5 secteurs clés : éducation, santé, travail, ville intelligente et police prédictive (avec une variante portant sur l'application des peines). Nous avons reçu un grand nombre de propositions qui ont été brièvement présentées dans la section précédente. Une année complète de coconstruction ne suffisant pas à couvrir tous les thèmes de réflexion, la démarche de la coconstruction se poursuivra à partir de septembre 2018 avec une série de consultations et d'ateliers de délibération au Québec, en Ontario et en Europe. Cette deuxième phase permettra d'aborder de nouveaux thèmes qui n'ont été qu'effleurés dans les scénarios utilisés lors de la première phase de coconstruction.

Parmi ceux-ci :

#### ENVIRONNEMENT

Les enjeux environnementaux sont, nous le savons tous, les enjeux vitaux que l'humanité doit affronter dans un avenir proche et lointain. L'IA peut aider à optimiser l'utilisation de nos ressources, mais elle est aussi gourmande en ressources et en énergie, en plus de générer des déchets électroniques. Le développement de l'IA est aussi un enjeu environnemental, mais l'avenir de la protection de l'environnement passera aussi par une utilisation ciblée de l'IA. Quels choix de société devons-nous faire pour répondre à nos responsabilités environnementales ?

#### DÉMOCRATIE ET PROPAGANDE DES MÉDIAS

L'IA et l'utilisation des mégadonnées présentent des défis majeurs pour la démocratie comme système de participation politique. Dès que l'on pense aux réseaux sociaux par exemple, viennent à l'esprit le scandale de l'utilisation des données des internautes à des fins politiques et l'intervention

de *chatbots* pour diffuser massivement de fausses nouvelles et contaminer un processus électoral. Notre rapport à l'intégrité des sources d'information et des médias est profondément affecté. Comment se prémunir contre un usage néfaste de l'IA dans le domaine politique ? Comment garantir les conditions d'exercice de l'autonomie critique des citoyens ?

#### SÉCURITÉ ET INTÉGRITÉ

Pour aborder les enjeux qui touchent au développement d'armes autonomes, de la police d'intervention à l'intérieur d'un pays (là où les recommandations ont le plus de chance d'avoir un impact), double-usage et mésusage de l'IA, intégrité des données (ex. protection des cyberattaques, etc.).

#### ARTS ET CULTURE

À l'ère des technologies numériques, on élabore, découvre, explore de nouvelles manières de produire des objets culturels et artistiques. L'impact de l'IA se manifeste à la fois sur la création artistique et sur la diffusion des œuvres. Une culture de l'IA se crée aussi, et notre rapport aux autres cultures s'en trouve modifié.

#### JUSTICE PRÉDICTIONNELLE

Si la question de la justice a été abordée dans le scénario sur la remise en liberté conditionnelle, il convient de l'approfondir en posant la question de l'avenir du jugement judiciaire et du droit, quand les algorithmes sont utilisés pour prédire la décision judiciaire. Ce domaine de développement algorithmique commence à bouleverser les pratiques des juges, des avocats et des médiateurs, mais il pourrait aussi modifier la manière dont est conçu le droit, en particulier la jurisprudence.

Des ateliers de coconstruction sur ces thèmes, sortiront un ensemble d'analyses et de propositions qui viendront compléter celles qui ont été produites lors de la première phase de coconstruction. Elles viendront aussi nourrir la réflexion sur les principes éthiques de la *Déclaration de Montréal* et sur les recommandations pour une politique publique de l'IA qui seront élaborées dans leur prolongement.

C'est autour de chantiers prioritaires que nous présenterons les recommandations de politiques publiques. Les chantiers prioritaires sont des axes de recommandations transversales aux secteurs et aux thèmes. Nous ne dévoilerons les chantiers prioritaires, et les recommandations, qu'une fois le processus de délibération terminé. Mais nous pouvons d'ores et déjà mentionner trois des chantiers qui se sont imposés :

- > Littérature numérique
- > Diversité et inclusion
- > Transition et mutations sociales

## 7.2

### UNE DÉMARCHE INSPIRANTE

Parmi les initiatives inspirées par la *Déclaration de Montréal*, il faut tout d'abord mentionner les travaux du groupe Montréal AI Ethics Meetup fondé et coordonné par Abhishek Gupta (McGill). Ce groupe qui rassemble plus d'une centaine de chercheurs multidisciplinaires et de citoyens concernés par les développements de l'IA, a consacré, entre décembre 2017 et mars 2018, plusieurs séances de 2 h aux principes de la Déclaration. S'il ne s'agit pas de délibérations citoyennes, les séances du Montréal AI Ethics Meetup sont néanmoins de véritables exercices d'intelligence collective, qui mobilisent une diversité de chercheurs de haut niveau. Un rapport détaillé issu de leurs réflexions critiques a été remis à l'équipe de la Déclaration et peut être consulté en ligne. Les auteurs du rapport sont : Stephanie Dyke, Paule-J Toussaint, Abhishek Gupta,

Gregory Caicos, Marc Daher, Peter Chen. Cette initiative est particulièrement encourageante car elle émane du cœur de la communauté montréalaise en IA.

Une autre initiative importante a été la soirée de réflexion organisée par l'ESG UQÀM (École des Sciences de la Gestion de l'UQÀM) le 15 février 2018, intitulée : « Vers un développement responsable de l'IA : Soirée de réflexion autour de la *Déclaration de Montréal* ». Cette soirée a réuni sept chercheurs de l'UQÀM autour des sept valeurs de la Déclaration. Après une présentation générale de Yoshua Bengio et Martin Gibert, le public a pu entendre les réflexions des professeurs Marie-Jean Meurs (bien-être), Christophe Malaterre (autonomie), Hugo Cyr (justice), Sébastien Gambis (vie privée), Étienne Harnad (connaissance), Dominic Martin (démocratie) et Maude Bonenfant (responsabilité). Le compte rendu des échanges qui ont eu lieu dans le cadre de cette rencontre universitaire est consultable en ligne<sup>22</sup>.

À l'Université de Montréal, la Faculté des Arts et des Sciences a créé en février 2018 Perspective, un laboratoire d'idées interdisciplinaire dont les premiers travaux s'inscrivent explicitement « dans le sillage de la *Déclaration de Montréal* pour un développement responsable de l'IA ». Ce laboratoire réunit un groupe d'étudiants aux cycles supérieurs chargés de produire des rapports dont l'objectif est d'éclairer les décideurs publics sur les impacts sociaux de l'IA.

Enfin, de nombreuses organisations (entreprises, organismes de développement ou associations) ont manifesté leur intérêt pour la *Déclaration de Montréal*, ont accueilli des présentations de la Déclaration ou organisé des discussions autour de ses principes. C'est le cas de IBM, Montréal InVivo, l'ACFAS, Printemps numérique, ou de C2Montréal.

Le processus de consultation et de participation citoyenne initié par la *Déclaration de Montréal* s'amplifie et se poursuit désormais hors du Québec, à Toronto sous l'impulsion de l'ICRA, et aussi prochainement en Europe, à Bruxelles et à Londres.

<sup>22</sup> [https://docs.wixstatic.com/ugd/ebc3a3\\_0e7d08f785c54b148d34c1c6c54f4b8c.pdf](https://docs.wixstatic.com/ugd/ebc3a3_0e7d08f785c54b148d34c1c6c54f4b8c.pdf)



# CONCLUSION

## Les nouvelles possibilités introduites par l'IA devraient profondément transformer la société dans les prochaines années. Mais comment s'assurer que ces algorithmes et jeux de données conduiront à des innovations sociales positives en santé, en éducation, dans le système judiciaire, dans les organisations publiques et privées, dans la vie quotidienne et dans nos villes ? Entre les nombreuses promesses créées par l'IA et les risques éthiques réels, comment se repérer ? Et comment en débattre ?

La démarche de coconstruction de la *Déclaration de Montréal* pour une IA responsable est d'abord née d'un constat : le débat sur l'IA responsable nous concerne tous et le sentiment d'incertitude et d'inquiétude que l'IA suscite est partagé. Ce débat ne peut donc pas être délégué aux seuls experts. À partir de ce constat, la démarche de coconstruction a relevé le pari qu'un débat associant les citoyens, les experts et les parties prenantes est tout à fait possible et même potentiellement générateur de propositions très innovantes, à condition de bien l'organiser.

À ce titre, ce rapport d'étape rédigé après plus de dix ateliers réalisés de février à mai 2018 à Montréal, Québec et Laval, présente un aperçu de deux résultats attendus à l'issue du processus de coconstruction en décembre 2018 : d'abord sur les processus mis en œuvre pour organiser la coconstruction, et puis sur les premières propositions thématiques formulées au cours de ces rencontres par les participants.

Sur les processus de participation déployés pour organiser ce débat, la formule du café citoyen de 3 h en bibliothèque publique, ouverte à tous, comme celle de la grande journée avec parties prenantes, citoyens et experts (à Montréal et Québec), ont toutes les deux généré des échanges très riches. En particulier, le choix de lieux publics inspirants pour tenir ces événements gratuits (bibliothèques publiques, Musée de la civilisation

à Québec, Société des arts technologiques à Montréal), l'attention apportée pour créer une convivialité entre participants, pour leur fournir des informations, des connaissances au bon moment (introduction pédagogique en début d'atelier et remise d'un livret du participant présentant la *Déclaration de Montréal* et des ressources pour la délibération en cours d'atelier), l'emploi de scénarios prospectifs présentant des scénarios d'usagers sur l'IA au Québec en 2025, permettant de donner aux participants des « ancrages » et des « déclencheurs » pour le débat tout en proposant des cadrages susceptibles d'éviter certains biais cognitifs, la facilitation des discussions par des facilitateurs réflexifs, partageant au fil des événements les manières d'animer ce type de délibération prospective, l'emploi d'affiches permettant de rythmer et de synthétiser les résultats intermédiaires des délibérations, tout cela a permis de mettre en place un dispositif d'atelier à la fois convivial, accueillant et générateur de propositions pertinentes et inédites dans un temps limité.

Ce type d'ateliers se poursuivra dans les prochains mois, avec l'idée d'expérimenter des variantes de dispositifs et de supports pour les débats. En particulier, le projet de développer un jeu de table sérieux pour stimuler la délibération prospective et éthique sur l'IA sera exploré et testé dans de prochains ateliers.

À l'issue de cette première étape de la délibération avec des citoyens, chercheurs et parties prenantes, de nombreux enjeux et pistes de solution ont été formulés par plusieurs centaines de citoyens qui se sont réunis autour des tables de coconstruction. L'objectif de la démarche étant de stimuler la délibération citoyenne sur le développement responsable de l'IA, les discussions se sont organisées à partir de scénarios mettant en scène des situations de prises de décision fictives, des enjeux éthiques, des risques possibles ou des controverses, pour à la fois exemplifier et mettre à l'épreuve les principes de la *Déclaration de Montréal* sur l'IA responsable.

Ces premiers résultats donnent une certaine idée de l'acceptabilité sociale de l'IA et de son développement. Pour les fins de cette analyse préliminaire, nous avons choisi de rester au plus

près de la voix des citoyens et des parties prenantes ayant participé aux débats. De plus, les résultats présentés à cette étape intermédiaire du processus de coconstruction (enjeux, pistes de solution, nouveaux principes potentiels) ne sont pas définitifs et seront explorés plus en profondeur lors des prochaines étapes de l'analyse, notamment sur les conditions de mise en œuvre.

Cette première phase de consultations a également mis en évidence certains dilemmes à explorer. Par exemple, dans le secteur de la santé, la protection de la confidentialité des données se confronte à la promesse d'une santé prédictive, préventive et personnalisée, qui nécessiterait pour les citoyens que l'IA prenne en considération une grande variété de données (et pas seulement des données biologiques). Autre exemple, dans les secteurs de la santé, de l'éducation, ou du système judiciaire, s'il y a accord pour dire que les humains doivent rester maîtres de la décision, cela peut se confronter dans les faits avec une volonté gestionnaire cherchant à automatiser certaines décisions pour faire des gains de productivité dans les organisations. Sur le plan des mesures à adopter, opter pour des mesures légalistes afin d'encadrer les entreprises peut aussi se confronter à la nécessité de soutenir des entreprises innovantes en croissance rapide qui pourraient contribuer demain à la prospérité d'un pays. Ces dilemmes sont des défis pour la gestion responsable de l'IA en société, et chacun peut être le point de départ de processus d'innovation collective pour imaginer des manières inédites de les surmonter.

Le développement de l'IA soulève ainsi de nombreuses questions éthiques et de société que le processus de coconstruction va continuer à analyser dans les prochains mois, en approfondissant les enjeux sectoriels explorés dans cette première étape, et en abordant de nouveaux thèmes : l'environnement et la transition énergétique, les liens entre démocratie, propagande et médias, la sécurité, les armes autonomes et l'intégrité des données, les arts et la culture. Toutes ces transformations portées par le développement de l'IA 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, ce sont les agencements entre humains et technologies qu'il est essentiel d'observer et d'analyser de manière prospective et critique. Si une revendication a fait consensus dans les débats de la coconstruction, c'est en effet celle de conserver une place centrale à l'humain dans un monde de plus en plus artificiellement intelligent.

# PARTICIPANTS À LA COCONSTRUCTION

**Merci aux étudiants et professionnels qui ont animé les ateliers et qui ont réalisé la prise de notes :**

**Alexandre Beaudoin-Peña**, Université de Montréal  
**Bhavish Beejan**, Université Laval  
**Karl Bherer**, Université Laval  
**Alexis Bibeau**, Université Laval  
**Pierre-Antoine Boutin-Panneton**, Université Laval  
**Dominic Cliche**, Université Laval  
**Eve Gaumont**, Université Laval  
**Emilie Guiraud**, Université Laval  
**Hubert Hamel-Lapointe**, Université de Montréal  
**Audrey Houle**, Université Laval  
**Nico Julien**, Université Laval

**Henri Lajeunesse**, Université Laval  
**Guillaume Macaux**, Université Laval  
**Mariève Mauger-Lavigne**, Université de Montréal  
**Orly Nahmias**, citoyenne  
**Judith Paquet**, Université Laval  
**Pierre-Luc Plante**, Université Laval  
**Lynda Robitaille**, directrice des opérations du Centre de recherche en données massives (CRDM)  
**Jason Stanley**, Université de Montréal  
**Yanis Taleb**, Université de Montréal  
**Clémence Varin**, Université Laval

**Merci aux citoyens, aux professionnels et aux experts ayant participé aux ateliers (seuls les noms des participants dont nous avons obtenu l'autorisation apparaissent ci-dessous) :**

Sihem Neila Abtroon  
Béatrice Alain  
Hassane Alami  
Rana Alvabi  
Alejandro Arreola-Alvarado  
Gabriel Arruda  
Barthélémy Aucourt  
Naomi Ayotte  
Manon Babine  
Maryluisa Barillas  
Philippe Beauchemin  
Stéphane Beaulieu  
François Beauregard  
Claude Bédard  
Sylvain Bédard  
Vincent Bergeron

Alexandre Berkesse  
Karl Bherer  
Marise Bonenfant  
Serge Bouchard  
Caroline Boudreault  
Lyne Bourbonnais  
Véronique Boutier  
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-Héту  
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	Jean-Michel Lapointe	Frédéric Plamondon	Daniel Tremblay
Julian Falardeau	Jonathan Lasprilla	Pier-Luc Plante	Jérémy Trudel
Simon Frappier	Sylvie Lavoie	Kamila Podgorska-	Marie-Christiane
Benoit Gagnon	Louis Lecaer	Gilbert	Trudel
Marie-Pierre Gagnon	Dominique Leclerc	Keith Poitras	Félix Vaillancourt
Marina Gallet	Sarah Legendre	Julie Politi	Julie Verdy
Hortense Gallois	Bilodeau	Thomas Poulin	Danael Villeneuve
Sébastien Gams	Pascale Lehoux	Denis Poussart	Grant Wark
Véronique Gareau-	Claude Lejeune	Emmanuelle Praise	Bryn Williams-Jones
Chiasson	Mélanie Levasseur	Louis-Philippe Pratte	Lemy Wong
Mathieu Gauthier-	Elisabeth Limoges	Mariel Ramos	William Wong
Pilote	Robert Locas	Diane Raymond	Almina Yagoubi
Sylvie Gélinas	Santiago Lopez	Catherine Régis	Ming Yue
Thomas George	Auréli Macé	Laurence Renault	
Gueno Gianni	Aicha Mafhoum	Cassie Rhéaume	
Jean-François Gignac	Suzanne Mainville	Toussaint Riendeau	
Martin Gibert	Mantas Manovas	Anne-Marie Robert	
Patricia Gingras	Mathieu Marcotte	François-Xavier	
Béatrice Godard	Jean-Pierre Marquis	Robert	
Christian Goudreau	Marie Martel	Louis-Nicolas Robert	
Gilles Gouin	Mariève Mauger-	Nicolas Roby	
Mervine Gowry	Lavigne	Stéphane Roche	
Alexandre Gravel	Moussa Mekhnach	Marie Roy	
Michel Grou	Natacha Mercure	Sara Russo-Garrido	
Alexandre Guédon	Bruno Milia	Laurence Sabourin	
Pascaline Guenou	Michael David Miller	Iger Sadoune	
François Guité	Ann Mitchell	Marie-Noëlle Saint-	
Carl Hamilton	Erica Monteferrante	Pierre	
Simon-Pierre Harvey	Farida Mostefaoui	James Sangster	
Lucie Hébert	Maria Moudfir	Sylvie Saucier	
Ghiles Helli	Vanessa Murray	Sébastien Adam	
Lucas Hubert	Orly Nahmias	Jean-François	
Aida Issa	Vanessa Nantel	Sénéchal	
Sabrina Jocelyn	John Newhouse	Eric Shannon	
Erwan Jonchères	Justin Ngoza	Danielle Sicotte	
Nico Julien	Zoonie Nguyen	Chantale Simard	
Debbie Jussome	Catherine Olivier	Julie Simard	
Ed Khazen	Daniel Pascot	Jean-Hébert Smith-	
Amy Khoury	Florence Paulhiac	Lacroix	
Andrée Labrie	Jorge Perez	Karima Smouk	
Anne-Marie Lacombe	Lorenzo Perozzi	Yanis Taleb	
Marie-Claude Lagacé	Geneviève Perreault	Isabelle Tanba	
Henri Lajeunesse	Benoit Petit	Christian Tanguay	
Karine Landry	Louis Piette	Marc Tomkinson	



# 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<sup>23</sup>.

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

Tableau 14 : 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	<b>Introduction pédagogique</b> Introduction aux enjeux éthiques et sociaux de l'intelligence artificielle ( <i>Déclaration de Montréal IA responsable</i> ), 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.

<sup>23</sup> 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 »)<sup>24</sup>.

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 : c'est, dit-il, « le futur interpelle le présent sur le mode de l'altérité »<sup>25</sup>. 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 inconnu, il devienne plus abordable, qu'il donne prise à la réflexion, à l'action »<sup>26</sup>.

Tableau 15 : 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	Introductions : principes de l'intelligence artificielle, les enjeux éthiques de l'IA ( <i>Déclaration de Montréal</i> ) 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 <i>Déclaration de Montréal</i> , 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

<sup>24</sup> Méthode du codesign développée par le Lab Ville Prospective de l'UdeM, [www.labvilleprospective.org](http://www.labvilleprospective.org)

<sup>25</sup> Michel De Certeau, *La culture au pluriel*, p. 223, Paris, Seuil, 1993

<sup>26</sup> Georges Amar, Prospective conceptive : pour un futur ouvert, revue *Futuribles*, 2015, n. 404, p. 21

## ANNEXE 2

### Les scénarios prospectifs

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 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 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 16 : 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.

Thème	Scénario sur l'IA en 2025	Résumé du scénario sur l'IA en 2025 au Québec
<b>2. Ville intelligente</b>	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 citoyens. 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 ( <i>nudges</i> )	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 ?
<b>3. Éducation prédictive</b>	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.
<b>4. Police prédictive et système judiciaire</b>	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).

Thème	Scénario sur l'IA en 2025	Résumé du scénario sur l'IA en 2025 au Québec
<b>5. Monde du travail</b>	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 police prédictive et le système judiciaire, 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'utilisateurs 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 17 : Constitution de cinq scénarios par thème

Scénarios sur l'IA en 2025	Jumeaux numériques	Voiture autonome	AlterEgo	Libération conditionnelle	Restructuration responsable
<b>Thèmes</b>	1. Santé prédictive	2. Ville intelligente	3. Éducation prédictive	4. Police prédictive et système judiciaire	5. Monde du travail
<b>Problématique sectorielle</b>	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
<b>Types d'apprentissage en IA</b>	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
<b>Enjeux éthiques et sociaux (exemples)</b>	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 : POLICE PRÉDICTIVE ET SYSTÈME JUDICIAIRE

### 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.<sup>27</sup> 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

<sup>27</sup> Une certification octroyée à des entreprises répondant à des exigences sociétales et environnementales, de gouvernance ainsi que de transparence envers le public.

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 co-bots ! 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'entraînement des algorithmes, et que ce n'est pas une solution respectable

pour les collègues. Il préférerait 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.

## ANNEXE 3

# Les autres formes de participation

Lors de sa mise en ligne en novembre 2017, le site web de la déclaration offrait deux possibilités de participation directe au processus de coconstruction : avec un questionnaire en ligne et une invitation à déposer un mémoire. Bien que les résultats de ces consultations seront davantage détaillés dans le rapport final, on peut d'ores et déjà en brosser ici un portrait des résultats préliminaires.

### A3.1 LE QUESTIONNAIRE EN LIGNE

Le questionnaire en ligne se composait de 35 questions, articulées autour des 7 valeurs de la version préliminaire de la *Déclaration de Montréal*. Il était bilingue. Un peu plus de 80 personnes ont répondu à des questions allant de « Comment l'IA peut-elle contribuer au bien-être ? » à « Un agent artificiel comme Tay, le *chatbot* « raciste » de Microsoft, peut-il être moralement blâmable et responsable ? », en passant par « Quels types de décisions de justice pourrait-on déléguer à une IA ? ». Ces questions donnent parfois lieu à des quasi-consensus : non, il n'est pas acceptable qu'une arme autonome tue un être humain ; oui, il faut lutter contre la concentration du pouvoir et de la richesse au sein d'un petit nombre d'entreprises en IA ; oui, on devrait savoir à qui nos données personnelles sont transmises et qui les utilisent. Mais ces questions suscitent également des doutes importants : l'IA peut-elle vraiment garantir le respect de la vie privée ? Est-il acceptable qu'une IA réponde en notre nom à des courriels ?

On voit aussi apparaître certains clivages, en particulier dans l'attitude à l'égard du secteur privé. Certains (la majorité) craignent une dérive de l'IA dictée par la recherche du profit des entreprises plutôt que par celle du bien commun (les voitures autonomes attisent les foudres de certains

répondants). D'autres considèrent le secteur privé comme la meilleure garantie qu'une IA se développe de façon indépendante si elle n'est pas liée à des programmes politiques.

Plus généralement, c'est probablement le niveau de confiance en l'IA (et en l'avenir) qui divise les gens de la façon la plus évidente. Alors que certains cherchent des solutions au problème suscité par l'IA en la rendant plus performante et adaptée au besoin humain, d'autres s'inquiètent des perturbations qu'elle va occasionner dans la vie sociale et économique, à la limite d'une déshumanisation.

Il est également remarquable que plusieurs pistes de solutions fassent écho à des recommandations concrètes apparues lors des consultations en bibliothèque ou durant les journées de coconstruction. On retrouve ainsi une valorisation de la transparence et l'idée de créer un ombudsman de l'IA ou un « comité de sages ». Enfin, au moins un leitmotiv ressort de cette consultation en ligne : c'est que l'IA doit être conçue comme un outil au service de l'humain. Comme l'écrit un répondant qui résume plusieurs interventions similaires, « les systèmes informatiques sont des aides à la décision et ils doivent le rester ».

## A3.2 LES MÉMOIRES OU RECOMMANDATIONS SPONTANÉES

En ce qui concerne les mémoires ou recommandations spontanées, 15 ont été reçus et sont disponibles sur le site de la *Déclaration de Montréal*<sup>28</sup>. Ceux-ci allant dans des directions très diverses, il est plus difficile d'en broser un portrait d'ensemble cohérent. Voici quelques réflexions ou suggestions qui s'en dégagent :

- > Une mise en garde contre le risque d'instrumentalisation et d'assimilation de l'humain à une simple machine qui dissimulerait une nouvelle idéologie totalitaire. « L'IA ne doit pas participer à la tentation de la déresponsabilisation humaine aux dépens de la technologie » (Jean-Claude Ravet, rédacteur en chef de la revue Relations).
- > Un appel à inclure les entreprises québécoises en technologie dans la réflexion sur le développement de l'IA (Association québécoise des technologies).
- > Un concept élaboré d'IA « loyale », c'est-à-dire que chacun d'entre nous posséderait une IA personnalisée, réglée pour agir au service exclusif de son propriétaire – et non au service d'une compagnie ou d'un État (G. Wark).
- > Quatre pistes pour concilier le développement de l'IA avec la protection de la vie privée (Commission d'accès à l'information du Québec).
- > La nécessité d'un organisme national ou international de surveillance et de recherche (un observatoire) qui proposerait des standards à respecter (John McNally).
- > La mise sur pied d'un organe consultatif indépendant composé de spécialistes en IA, éthique et droit, et de citoyens (Lise Parent).
- > Un argument pour que le partage d'information ne soit pas l'option par défaut pour ce qui concerne le principe de vie privée, et une mise

en garde contre une dérive technocratique et non réellement démocratique de la *Déclaration de Montréal* (Ariane Quintal, Matthew Sample et Eric Racine).

- > Un argumentaire pour une conception moralement minimaliste de l'IA (Pierre Musseau-Milesi).
- > Enfin, l'Ordre des ingénieurs du Québec note que les valeurs de la Déclaration sont « parfaitement compatibles avec les valeurs de la profession d'ingénieur » et propose plusieurs recommandations liées à la formation et à la diversité ainsi que l'adoption de « bonnes pratiques » comme, par exemple, le Règlement européen sur la protection des données (RGPD).

Nous tenons par ailleurs à remercier tous les citoyens et organisations ayant contribué par écrit à enrichir la réflexion sur le développement responsable de l'IA :

- > L'Ordre des ingénieurs du Québec
- > Bruno Robert
- > Jean-Claude Ravet
- > L'Association québécoise des technologies
- > Annick, Guillaume et Raphaël Hernandez
- > G. Wark
- > La Commission d'accès à l'information
- > John McNally
- > L'Université du Québec à Montréal
- > Lise Parent
- > Ariane Quintal, Matthew Sample et Éric Racine
- > Pierre Musseau-Milesi
- > Human Aware
- > Et ceux et celles qui ont souhaité rester anonymes.

L'ensemble des mémoires est disponible sur notre site web<sup>28</sup>.

<sup>28</sup> [www.declarationmontreal-iaresponsable.com/propositions-citoyennes](http://www.declarationmontreal-iaresponsable.com/propositions-citoyennes)

