



< >

Déclaration de Montréal

IA responsable_

</ >

PARTIE 1

DÉMARCHE ET MÉTHODOLOGIE

This part of the declaration is composed of several pages of extremely dense and illegible text, appearing as a single, continuous stream of characters. The text is primarily in black and yellow, with some red and blue highlights. It discusses the methodology and approach taken for responsible AI development, likely detailing the ethical framework, research methods, and operational practices used.

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

RÉDACTION

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

CHRISTOPHE ABRASSART, codirecteur scientifique de la coconstruction, professeur à l'École de design et codirecteur du Lab Ville Prospective à la Faculté de l'aménagement de l'Université de Montréal, membre du Centre de recherche en éthique (CRÉ)

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

TABLE DES MATIÈRES

| | |
|--|-----------|
| SOMMAIRE | 25 |
| 1. INTRODUCTION | 28 |
| 2. POURQUOI LA DÉCLARATION DE MONTRÉAL IA RESPONSABLE? | 31 |
| 2.1 L'origine intellectuelle du projet | 32 |
| 2.2 Le Forum sur le développement socialement responsable de l'intelligence artificielle | 34 |
| 2.3 Vers la Déclaration de Montréal pour un développement responsable de l'IA | 35 |
| 2.4 Montréal et le contexte international | 36 |
| 3. LES ENJEUX ÉTHIQUES ET SOCIÉTAUX DE L'IA | 39 |
| 3.1 Se faire une idée de l'IA | 39 |
| 3.2 L'IA au quotidien et le questionnement philosophique | 41 |
| 3.3 Les enjeux éthiques de l'IA | 43 |
| 3.4 L'éthique de l'IA et la Déclaration de Montréal | 45 |
| 4. LA DÉMARCHE DE COCONSTRUCTION | 47 |
| 4.1 Les principes de la démarche de coconstruction | 47 |
| 4.1.1 Les principes d'une bonne participation citoyenne | 47 |
| 4.1.2 Experts et citoyens | 49 |
| 4.2 La méthodologie des ateliers de coconstruction | 50 |
| 4.3 Originalité de la démarche de coconstruction | 52 |
| 4.4 Cafés citoyens en marge des bibliothèques | 53 |
| 4.5 Portrait des participants | 53 |

TABLE DES MATIÈRES

| | |
|--|-----------|
| 5. PARCOURS DÉLIBÉRATIFS DANS LES ATELIERS : EXEMPLES DE DEUX SECTEURS : VILLE INTELLIGENTE ET MONDE DU TRAVAIL | 56 |
| 5.1 Les parcours délibératifs | 56 |
| 5.1.1 Secteur ville intelligente : la voiture autonome (VA) et le juste partage de la rue | 57 |
| 5.1.2 Secteur du monde du travail : une restructuration socialement responsable? | 61 |
| 6. PARTICIPANTS À LA COCONSTRUCTION ET ÉQUipes DE TRAVAIL | 66 |
| ANNEXES | 71 |
| Annexe 1 | |
| Les ateliers de coconstruction : description détaillée et fonctionnement | 71 |
| Les cafés citoyens | 71 |
| Les journées de coconstruction | 72 |
| Annexe 2 | |
| Les scénarios prospectifs de la coconstruction de l'hiver | 74 |
| CRÉDITS | I |
| PARTENAIRES | II |

TABLE DES TABLEAUX ET DES FIGURES

| | |
|--|----|
| Figure 1 : Les valeurs de la Déclaration (version préliminaire) | 31 |
| Figure 2 : Les valeurs de la Déclaration de Montréal IA responsable | 32 |
| Figure 3 : La démarche de coconstruction | 35 |
| Figure 4 : La prospective stratégique : une démarche en trois temps | 52 |
| Figure 5 : Proportion hommes-femmes ayant participé aux ateliers de coconstruction | 54 |
| Figure 6 : Les participants aux ateliers de coconstruction par tranches d'âge | 54 |
| Figure 7 : Répartition des répondants aux cafés citoyens et aux journées de coconstruction par niveau de scolarité atteint | 54 |
| Figure 8 : Répartition des répondants aux cafés citoyens et aux journées de coconstruction par secteur d'activité | 55 |
| Tableau 1 : Ville intelligente, Premier moment délibératif: formulation d'enjeux éthiques en 2025 | 58 |
| Tableau 2 : Ville intelligente, Deuxième moment délibératif: propositions d'encadrement de l'IA pour 2018-2020 | 59 |
| Tableau 3: Monde du travail, Premier moment délibératif: formulation d'enjeux éthiques en 2025 | 63 |
| Tableau 4 : Monde du travail, Deuxième moment délibératif: proposition d'encadrement de l'IA pour 2018-2020 | 64 |
| Tableau 5 : Déroulement type des cafés citoyens | 71 |
| Tableau 6 : Déroulement type des journées de coconstruction | 72 |
| Tableau 7 : Résumé des scénarios | 74 |
| Tableau 8 : Constitution de cinq scénarios par thème | 76 |

SOMMAIRE

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

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

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

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

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

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

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

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

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

Les résultats des ateliers de coconstruction – Les grandes orientations

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Les mécanismes pour la transition numérique

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

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

Poursuivre la délibération

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

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

1. INTRODUCTION

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

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

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

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

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

De nombreuses objections et craintes ont été exprimées lors de ce premier parcours de coconstruction. Plusieurs participants aux ateliers et plusieurs observateurs des travaux de la Déclaration ont mis en cause l'idéologie techniciste qui voit dans la technologie le moyen d'organiser rationnellement toute la société et qui réduit les enjeux sociaux à des problèmes techniques. D'autres questionnent la capacité et la volonté des institutions publiques à réguler des technologies lucratives. Ces objections ne doivent pas être balayées d'un revers de main parce qu'elles sont fondées sur des précédents historiques qui ont ébranlé la confiance dans les innovations technologiques et plus encore dans leurs promoteurs. Mais il est important que ceux

qui formulent ces objections fassent aussi en sorte de ne pas miner tout effort pour orienter positivement l'avenir de notre société et soutiennent, en y participant, la délibération démocratique qui nous permet de garder le contrôle sur le développement des technologies du numérique et de l'intelligence artificielle. On peut déplorer les effets de ces nouvelles technologies sur le lien social, on peut critiquer la réduction de la vie sociale à un ensemble de modes de vie, cela n'arrêtera pas l'innovation technologique, ni ne la fera dévier. Or, c'est tout l'enjeu de la *Déclaration de Montréal* : orienter le développement de l'intelligence artificielle afin de promouvoir des intérêts éthiques et sociaux fondamentaux et offrir des repères pour protéger les droits humains.

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

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

Si c'était le cas, il faudrait aussi renoncer à parler de bras mécanique au prétexte qu'un bras biologique est très différent dans son fonctionnement, et que les os, les articulations, les tendons et les muscles ne sont pas des pièces de métal, des poulies, des ressorts et des cordes. Ensuite, on confond souvent l'intelligence et la pensée en général. L'intelligence est une propriété de la pensée, elle n'est pas toute la pensée. L'intelligence a ceci de particulier qu'elle réduit la complexité du monde dans lequel l'être intelligent évolue pour lui permettre de mieux maîtriser son environnement. On se donne des règles pour analyser la réalité, calculer, évaluer et prendre des décisions. Une longue tradition philosophique de penseurs qui ne manquaient pas d'intelligence, n'a cessé de l'affirmer depuis Socrate à Russell en passant par Leibniz. D'une certaine manière, l'intelligence modélise et appauvrit la réalité pour mieux agir dessus, comme une équation en mécanique modélise et appauvrit le mouvement pour mieux le saisir. Enfin, et cela découle de ce qui précède, l'intelligence, même humaine, est dans une large mesure algorithmique : elle analyse des données et calcule selon des procédures. Elle se prête alors très bien à une « mécanisation » et à une « incarnation » au sens littéral du terme : le calcul digital, c'est-à-dire le calcul avec et sur les doigts selon des techniques très variées, est une incarnation du calcul ; avec les différents abaque comme le boulier chinois, la Pascaline¹ et la calculatrice électronique, on assiste à une mécanisation du calcul.

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

d'une intelligence artificielle. Dans le monde que nous connaissons et dans celui que nous pouvons anticiper à court et moyen terme, la réflexion sur les finalités de la vie sociale et de l'existence en général est le produit de l'intelligence humaine.

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

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

2. POURQUOI LA DÉCLARATION DE MONTRÉAL IA RESPONSABLE ?

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

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

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

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

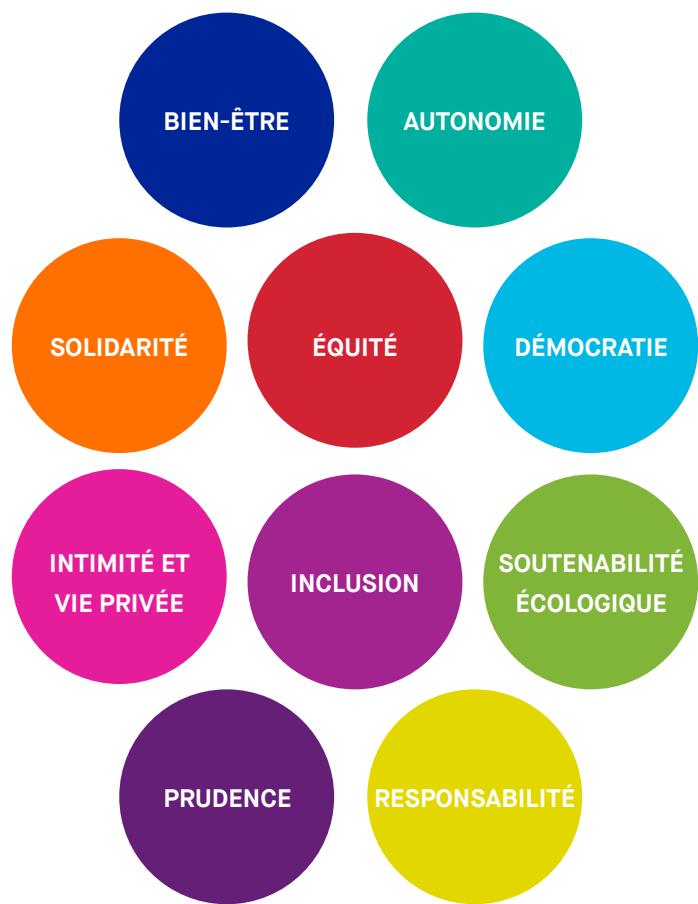
Figure 1: Les valeurs de la Déclaration (version préliminaire)



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

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

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



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

2.1

L'ORIGINE INTELLECTUELLE DU PROJET

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

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

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

L'utilisation d'algorithmes pour prendre des décisions techniques ou administratives n'est pas nouvelle. Si les algorithmes sont connus depuis le Moyen Âge², 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, prédition d'accident cardiaque³. Toutes ces techniques algorithmiques, « les procédures », posent déjà un certain nombre d'enjeux éthiques et sociaux : celui de l'acceptabilité sociale de la décision « automatique », celui de la dernière décision (un humain est-il au bout de la chaîne de décision ?), ou encore de la responsabilité en cas d'erreur. Et il est évident que ces enjeux se posent de nouveau avec les dernières innovations algorithmiques.

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

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

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

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

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

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

2.2

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

Ces réflexions ont été le point de départ de l'initiative des Fonds de recherche du Québec et de l'Université de Montréal pour organiser une rencontre internationale faisant le point sur les impacts sociaux de l'IA. Dans ce cadre, le comité d'organisation à l'Université de Montréal proposait de lancer les travaux de la *Déclaration de Montréal pour un développement responsable de l'IA* sur la base d'un processus de consultation et de participation⁴. Les 2 et 3 novembre 2017, s'est ainsi tenu au Palais des congrès de Montréal, un forum rassemblant les plus grands experts des domaines concernés par la réflexion sur l'IA, des sciences fondamentales aux sciences humaines et sociales.

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

- > Offrir un espace de réflexion public quant aux enjeux du développement de l'IA et ses impacts sociaux
- > Intéresser et sensibiliser les décideurs, les partenaires industriels, les représentants politiques et la communauté qui s'intéressent à l'IA quant aux questions de société soulevées par son essor et ses applications
- > Valoriser une approche interdisciplinaire et intersectorielle comme facteur de réussite essentiel au développement éthique et durable de l'IA

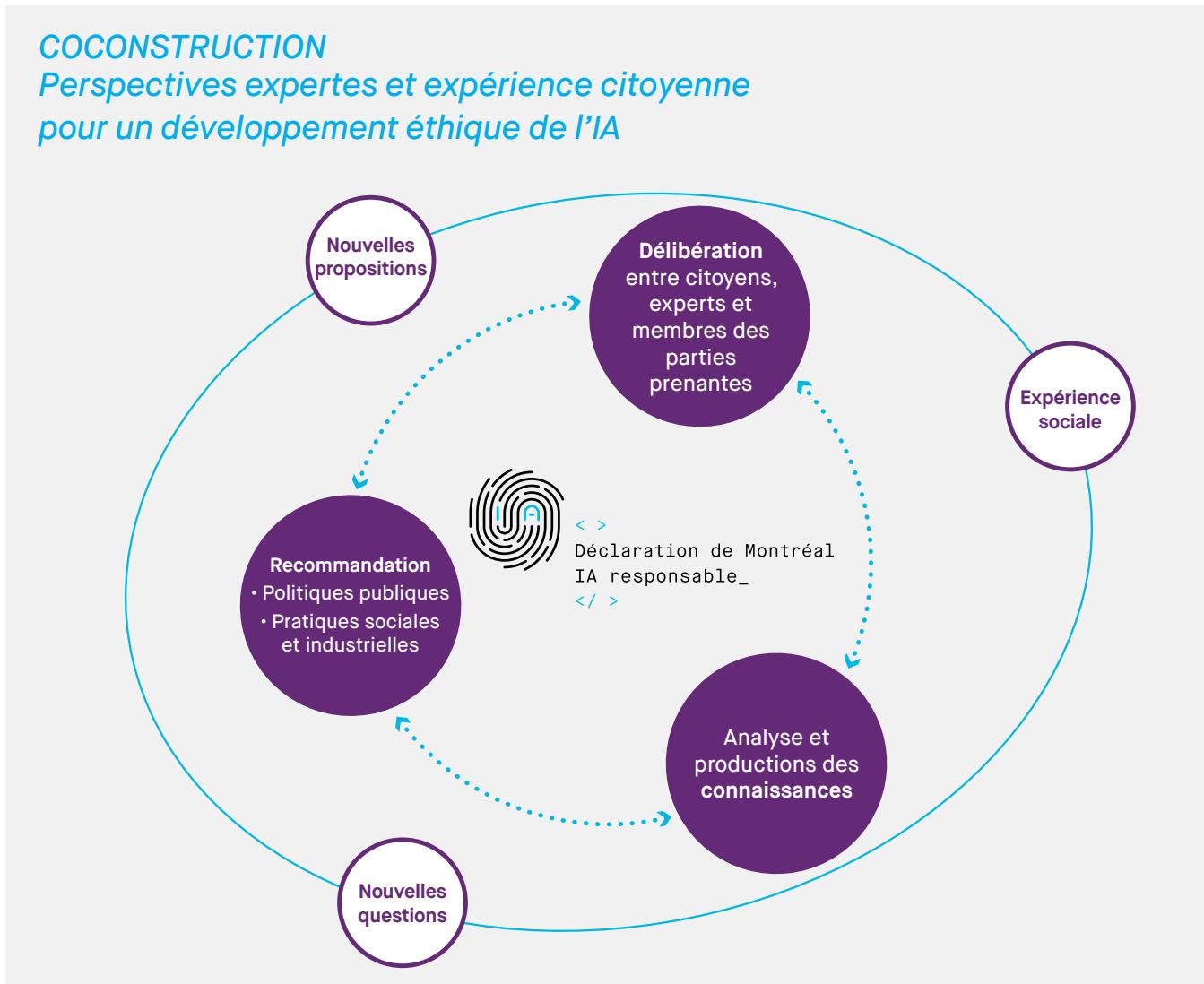
Se sont ainsi définis les contours d'une démarche inclusive (interdisciplinaire et intersectorielle) qui est au cœur de l'entreprise d'élaboration de la *Déclaration de Montréal* pour un développement de l'IA qui soit à la fois responsable, vecteur de progrès social et garant de l'égalité et de la justice.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

⁴ Le comité scientifique du Forum était composé de Louise Béliveau (UdeM, Vice-rectorat aux affaires étudiantes et aux études), Yoshua Bengio (UdeM, Département d'informatique, MILA, IVADO), David Décaray-Hétu (UdeM, École de criminologie), Nathalie De Marcellis-Warin (École Polytechnique, Département de mathématiques et de génie industriel, CIRANO – Centre interuniversitaire de recherche en analyse des 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).

2.3

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

Figure 3 : La démarche de coconstruction



Comme nous l'avons mentionné en début de chapitre, le travail initial d'identification de ces valeurs et des principes correspondants avait pour seul objectif de lancer un processus de participation citoyenne afin de préciser les principes éthiques d'un développement responsable de l'IA, de les enrichir et de les compléter. On ne s'étonnera donc

pas que la version préliminaire de la Déclaration soit schématique et que l'énoncé des principes soit volontairement très simple et consensuel, laissant à la délibération publique la possibilité de les interpréter et de les compléter⁵. Un an plus tard, la Déclaration a été substantiellement enrichie.

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

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

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

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

recommandations d'encadrement éthique et de politique publique, devra se prolonger dans une structure institutionnelle pérenne qui permette de rester réactif face aux évolutions de l'IA.

2.4

MONTRÉAL ET LE CONTEXTE INTERNATIONAL

L'initiative de la *Déclaration de Montréal* s'inscrit dans un contexte scientifique, social et industriel favorable. Montréal est un pôle de recherche et de développement majeur en intelligence artificielle avec une communauté de chercheurs et des laboratoires universitaires de réputation mondiale (MILA, IVADO) et une pépinière de start-ups et d'entreprises en plein essor. Ce développement scientifique, technologique et industriel est au cœur d'une révolution des pratiques sociales, des modèles économiques et des modes de vie, qui touche tous les secteurs de la société. La Ville de Montréal, avec son Laboratoire de l'innovation urbaine⁶, est aussi un laboratoire vivant du changement social et technologique. Avec la recherche scientifique fondamentale viennent des responsabilités éthiques et sociales que la communauté montréalaise de l'IA assume pleinement.

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

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

Plusieurs interlocuteurs internationaux ont également manifesté leur intérêt pour la *Déclaration de Montréal*, notamment pour sa méthode d'élaboration. L'équipe de la Déclaration a pu établir un dialogue avec des institutions comme la Royal Society du Royaume-Uni⁷, l'EGE (*European Group on Ethics in Science and New Technologies*⁸) et le HLEG (*High Level Expert Group on AI*⁹) de la Commission européenne qui ont leur propre programme d'étude et de recommandation sur l'IA. On note d'abord une convergence des lignes directrices pour un développement éthique de l'IA ainsi qu'une volonté partagée de faire valoir une conception démocratique de l'utilisation de l'IA au service du bien commun.

La démarche de la *Déclaration de Montréal* doit ainsi se comprendre dans le contexte international d'un **printemps de l'IA**. Elle est précédée par plusieurs initiatives qui doivent être saluées car elles ont catalysé la réflexion sur une IA responsable. Il faut tout d'abord évoquer la création en 2014 du *Future of Life Institute* qui a produit en 2017 la Déclaration d'Asilomar : à l'issue d'une conférence de 3 jours, une déclaration contenant 23 principes fondamentaux encadrant la recherche sur l'IA et ses applications a été signée par plus de 1200 chercheurs. Y participait alors le professeur Yoshua Bengio qui attirait l'attention sur les risques d'utilisation irresponsable et malveillante de l'IA¹⁰.

Depuis la conférence d'Asilomar, plusieurs rapports sur l'éthique de l'IA ont été publiés. Le rapport de l'Association internationale des ingénieurs électriques 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'IA¹¹. Deux rapports stratégiques ambitieux ont été publiés en mars et avril 2018 : le rapport de la Mission Villani en France et celui de la Chambre des lords au Royaume-Uni « *AI in the UK : ready, willing, and able?* ». Il faut aussi souligner la démarche participative de la CNIL (Commission nationale de l'informatique et des libertés) en France qui a débouché sur la publication du rapport au titre évocateur : « *Comment permettre à l'Homme de garder la main ? – Les enjeux éthiques des algorithmes et de l'intelligence artificielle* », en décembre 2017.

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

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

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

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

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

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

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

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

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

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

3. LES ENJEUX ÉTHIQUES ET SOCIÉTAUX DE L'IA

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

3.1

SE FAIRE UNE IDÉE DE L'IA

L'idée de l'IA n'est pas nouvelle. Il faudrait au moins remonter au 17^e 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 algorithmique¹³. La notion de *calculus ratiocinator* (le calcul logique) préfigure l'idée de machine intelligente telle qu'elle sera développée trois siècles plus tard, dans les années 1940, par Alan Turing. En 1948, dans un rapport intitulé « *Intelligent Machinery* » et en 1950, dans son fameux article « *Computing Machinery and Intelligence*¹⁴ », Alan Turing évoque l'intelligence de la machine et élaboré 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^e siècle pour assister au printemps de l'IA.

L'IA consiste d'une certaine manière à simuler l'intelligence humaine¹⁵, s'en inspirer et la reproduire. Mais dans un premier temps, c'est le cerveau, siège de l'intelligence humaine, qui a été conçu comme une machine capable de recueillir, percevoir, et collecter des données de son environnement qu'il va ensuite analyser, interpréter et comprendre, se nourrissant de ces expériences pour établir des relations. Le domaine de recherche de l'IA consiste à produire des outils mathématiques pour formaliser les opérations de l'esprit et ainsi créer des machines qui peuvent accomplir des tâches cognitives plus ou moins générales, associées à l'intelligence humaine naturelle. Par exemple, découvrir des motifs complexes parmi une grande quantité de données, ou encore raisonner de manière probabiliste, afin de classer en fonction de catégories des informations, de prédire une donnée quantitative ou de regrouper des données ensemble. Ces compétences cognitives sont à la base d'autres compétences comme celles de décider entre plusieurs actions possibles pour réaliser un objectif, d'interpréter une image ou un son, de prédire un comportement, d'anticiper un événement, de diagnostiquer une pathologie, etc.

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

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

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

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

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

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

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

Mais ce qui a changé, c'est aussi le type d'algorithme élaboré par les chercheurs en IA. Les algorithmes déterministes, qui sont une suite déterminée d'instructions comme une recette de cuisine, laissent désormais la place à des algorithmes apprenants et reposent sur des réseaux neuronaux de plus en plus complexes à mesure que la puissance de calcul des machines augmente. En informatique, on parle de *machine learning* (apprentissage machine ou apprentissage

automatique) et les progrès de ce secteur de recherche ont été renforcés par le développement du *deep learning* (apprentissage profond). Au cœur de la notion même d'IA, se trouve la capacité d'adaptation et d'apprentissage. En effet, pour qu'une machine puisse être considérée comme intelligente, il faut qu'elle soit capable d'apprendre par elle-même à partir des données qui la nourrissent, comme le fait un être humain. Et comme pour l'être humain, l'apprentissage machine peut être supervisé, ou non supervisé, par des êtres humains qui entraînent les machines sur les données.

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

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

3.2

L'IA AU QUOTIDIEN ET LE QUESTIONNEMENT PHILOSOPHIQUE

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

Ces technologies sont si bien intégrées dans notre quotidien que nous n'y pensons plus vraiment. Quand on évoque l'IA, la plupart de gens l'associent encore à des machines menaçantes, polyvalentes et dotées d'une forme de conscience, capables de former un plan pour se débarrasser des êtres humains¹⁶. 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, 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 ? »¹⁷. Or, dès l'introduction de son article, il change le problème auquel il estime pouvoir donner une solution : une machine peut-elle se comporter de telle sorte qu'on ne puisse pas faire la différence avec une personne humaine ? Il propose alors le fameux « jeu de l'imitation » qui consiste à mettre en communication un être humain qui pose des questions (l'interrogateur) avec un autre être humain et une machine qui répondent à ses questions. Si la machine imite assez bien l'être humain au point que l'interrogateur ne parvient pas à dire qui de l'être humain ou de la machine a répondu, nous pouvons considérer que la machine pense. C'est cela qu'on désigne par l'expression « test de Turing ».

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

¹⁷ A. M. Turing (1950).

Ce jeu de l'imitation a fait couler beaucoup d'encre et les philosophes se sont durement opposés les uns aux autres pour savoir si nous pouvions dire qu'une machine pense. Une expérience connue sous le nom de « la chambre chinoise » a été popularisée dans les années 1980 par le philosophe John Searle¹⁸. 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 d'un être artificiellement

intelligent, les machines intelligentes restent des outils développés pour accomplir des tâches bien définies. On peut donc laisser à la philosophie cognitive, à la métaphysique, la psychologie et aux neurosciences le soin de débattre de la notion d'intelligence artificielle et discuter de la possibilité que les robots développent des émotions et éprouvent de l'empathie¹⁹. Le questionnement que pose l'introduction des IA dans nos vies est de type pratique, qu'il soit éthique, politique ou juridique. C'est un questionnement d'abord sur les valeurs et les principes éthiques, sur les orientations des politiques publiques et sur l'application de normes pour encadrer la recherche en IA et ses applications.

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

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

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

3.3

LES ENJEUX ÉTHIQUES DE L'IA

Pourquoi introduire l'éthique quand on aborde les impacts sociaux, 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 sociaux 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 contrôlée²⁰? 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.

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

3.4

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

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

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

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

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

2. Bien qu'ils soient divers, ils doivent faire l'objet d'une interprétation cohérente afin d'éviter tout conflit qui empêche leur application. D'une manière générale, les limites de l'application d'un principe sont tracées par le domaine d'application d'un autre principe.
3. Bien qu'ils reflètent la culture morale et politique de la société dans laquelle ils ont été élaborés, ils constituent une base pour une dialogue interculturel et international.
4. Bien qu'ils puissent être interprétés de diverses manières, ils ne peuvent pas être interprétés de n'importe quelle manière. Il est impératif que l'interprétation soit cohérente.
5. Bien que ce soient des principes éthiques, ils peuvent être traduits en langage politique et interprétés de manière juridique.

La Déclaration de principes est suivie d'une liste de recommandations dont l'objectif est de proposer des lignes directrices pour réaliser la transition numérique dans le cadre éthique de la Déclaration. Cette liste n'a pas vocation à être exhaustive et ne peut couvrir tous les secteurs d'application de l'IA ; ce serait une ambition vouée à l'échec. Il s'agit plutôt de couvrir quelques thèmes intersectoriels clés pour penser la transition vers une société dans laquelle l'IA permet de promouvoir le bien commun : la gouvernance algorithmique, la littératie numérique, l'inclusion numérique de la diversité et la soutenabilité écologique.

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

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

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

4. LA DÉMARCHE DE COCONSTRUCTION

4.1

LES PRINCIPES DE LA DÉMARCHE DE COCONSTRUCTION

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

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

C'est tout le sens de la démarche initiée par la Déclaration de Montréal : rendre à la démocratie la compétence de trancher les questions morales

et politiques qui concernent la société dans son ensemble. L'avenir de l'IA n'est pas seulement écrit dans des algorithmes, il réside d'abord dans l'intelligence humaine collective.

4.1.1 Les principes d'une bonne participation citoyenne

Dès lors que l'on fait intervenir le public dans un processus de consultation et de participation sur des problématiques sociales controversées, il faut s'assurer que ce processus soit conduit de façon à éviter les risques habituellement associés à l'exercice démocratique. On fait traditionnellement deux objections pour disqualifier le recours au public²¹:

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 tyanniques à l'égard des minorités.*

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

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

A. LA DIVERSITÉ ÉPISTÉMIQUE

Il faut tout d'abord s'assurer que les groupes délibérants manifestent la plus grande diversité interne, en termes de milieu social, de genre, de génération, ou d'origine ethnique. Cette diversité n'est pas seulement requise par l'idée que l'on se fait d'une démocratie inclusive, elle est aussi nécessaire pour augmenter la qualité épistémique des débats. Cela signifie simplement que chacun apporte une perspective différente sur le sujet débattu et enrichit la discussion²².

B. L'ACCÈS À UNE INFORMATION PERTINENTE

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

C. LA MODÉRATION

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

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

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

D. L'ITÉRATION

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

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

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

Or, dans les très grands groupes, on ne peut pas s'assurer que tous les individus aient la compétence requise et que chacun ait plus d'une chance sur deux d'avoir une opinion adéquate. Permettre la délibération (la communication entre eux) est un des moyens d'augmenter la compétence des participants pourvu qu'elle soit encadrée, comme nous l'avons fait. Certes, cela ne satisfait pas la 2^e condition de Condorcet, mais cela permet de garantir la 1^{re} 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 participation²³.

Pour toutes ces raisons, nous avons privilégié la forme de l'atelier de coconstruction qui réunit des citoyens non experts, des experts, des parties prenantes (associations, syndicats, représentants professionnels, entreprises) et des acteurs de la vie politique. Ces ateliers sont organisés selon des formats différents qui sont adaptés aux lieux de délibération et aux disponibilités des participants, et permettent de satisfaire aux conditions d'une participation citoyenne féconde et robuste. Mais il faut noter que le processus d'élaboration de la Déclaration est complexe et repose sur d'autres

types de consultations : questionnaires en ligne, rapports et tables rondes d'experts. La Déclaration n'est pas l'enregistrement pur et simple de la parole recueillie dans les ateliers de coconstruction ; elle est le fruit d'une délibération multiple et d'une réflexion qui s'appuie sur les ateliers de coconstruction.

4.1.2 Experts et citoyens

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

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

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

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

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

- > **S'assurer des conditions favorables de la participation citoyenne**
- > **Clarifier les enjeux éthiques qui sous-tendent les controverses autour de l'intelligence artificielle**
- > **Rationaliser les arguments défendus par les participants en leur indiquant les arguments que l'on sait erronés ou biaisés et en leur expliquant les raisons pour lesquelles ils sont erronés.**

Le rôle des éthiciens est donc celui d'un accompagnement éclairé²⁴. Les experts dans les autres domaines de recherche (en informatique, en santé, en sécurité, en droit, etc.) ont également un rôle d'accompagnement en fournissant aux participants les informations les plus utiles et les plus fiables sur l'objet de la controverse : Comment fonctionne un algorithme qui apprend à établir des diagnostics ? Le médecin peut-il être remplacé par un robot programmé pour le diagnostic ? Quelles sont les protections que nous pouvons opposer aux tentatives de piratage de nos données médicales ? Etc.

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

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

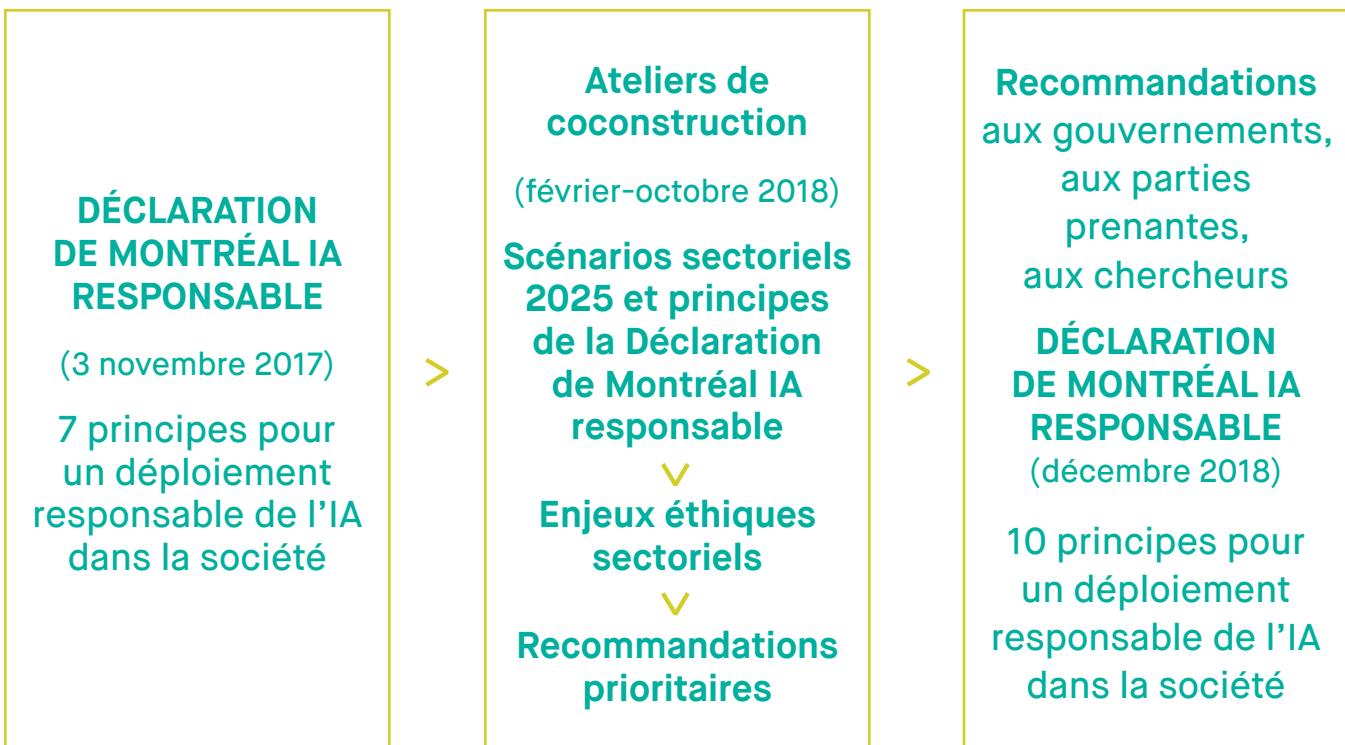
4.2

LA MÉTHODOLOGIE DES ATELIERS DE COCONSTRUCTION

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

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

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



Plus de 10 ateliers de coconstruction ont été organisés de février à octobre : des cafés citoyens de 3 heures dans des bibliothèques publiques, et deux grandes journées de coconstruction avec des citoyens, des experts et des parties prenantes variées (à la SAT à Montréal, au Musée de la civilisation à Québec et enfin au Centre culturel canadien²⁵).

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 Canada²⁶. En passant du modèle de l'espace de prêt documentaire à celui de la « bibliothèque lieu » 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.

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

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

4.3

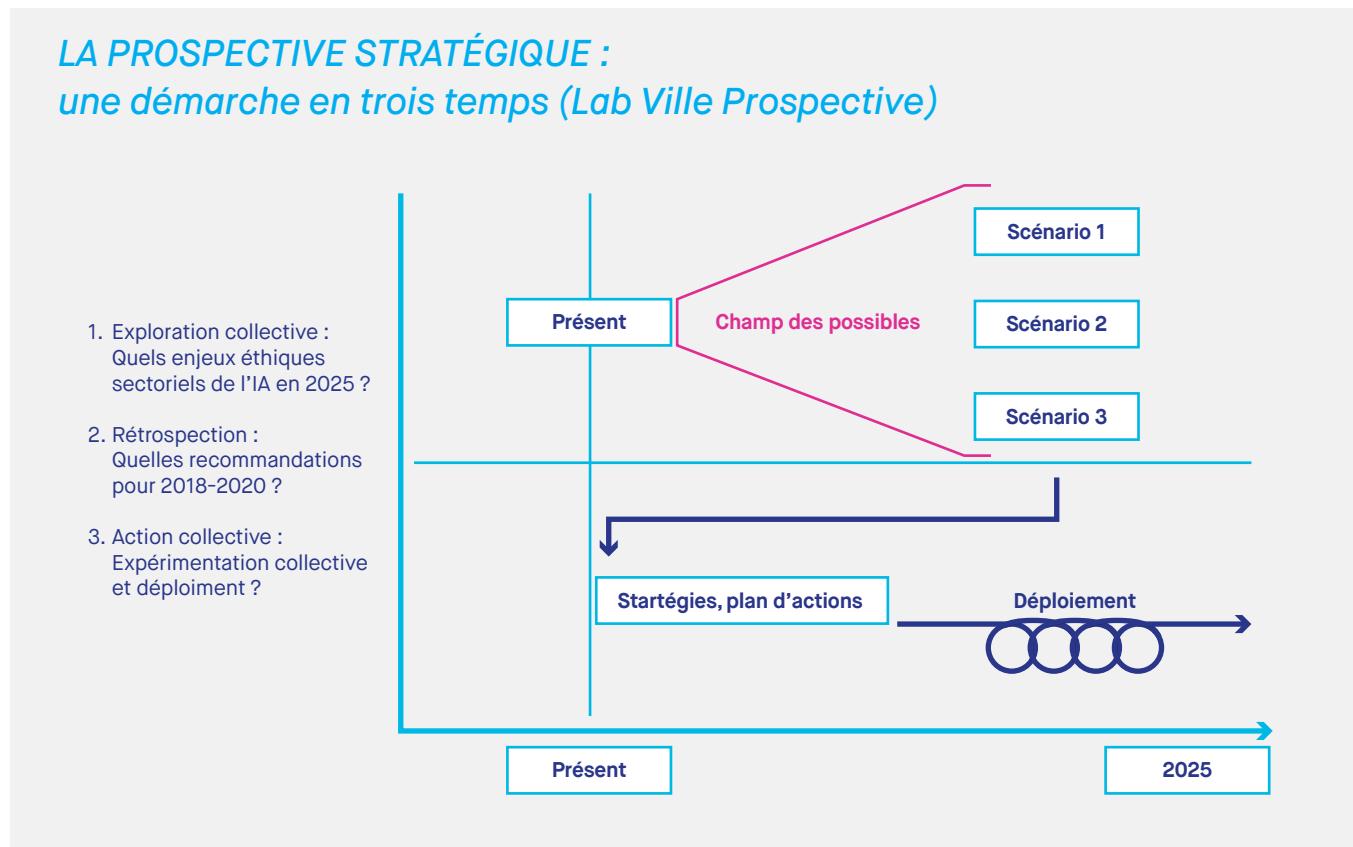
ORIGINALITÉ DE LA DÉMARCHE DE COCONSTRUCTION

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

- > Tout d'abord, le recours aux méthodes de prospective stratégique (« foresight »), avec des scénarios sectoriels en 2025 exemplifiant par de courts récits comment des controverses éthiques sur l'IA pourraient surgir ou s'amplifier dans les prochaines années (dans les secteurs de la santé, de la justice, de la ville intelligente,

de l'éducation et de la culture, du monde du travail). Ces scénarios 2025, qui présentent une variété de situations possibles face à un avenir très ouvert, seront utilisés comme déclencheurs de débats, pour identifier, préciser ou anticiper des enjeux éthiques sectoriels sur le déploiement de l'IA dans les prochaines années. Ces discussions à l'horizon 2025 permettront ensuite de formuler rétrospectivement des recommandations concrètes pour 2018-2020, pour nous diriger vers des situations collectivement souhaitables.

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



- > Ensuite, le recours à des méthodes d'animation de design participatif en « forum hybride »²⁷ pluridisciplinaire, incluant les citoyens et les parties prenantes, dans un contexte d'incertitude partagée face aux futurs possibles (pour approfondir un scénario, concevoir des dispositifs de réponse à un risque éthique, proposer un complément à la Déclaration de Montréal en cas d'enjeu orphelin, i.e. sans principe éthique correspondant).
- > Enfin, une attention aux « biais de paradigmes » qui ont des effets de cadrage très puissants dans la manière de poser les problèmes (ex. aborder les enjeux éthiques de la voiture autonome uniquement sous l'angle du dilemme du tramway comme le propose l'équipe de l'expérience *Moral Machine* du MIT) et dans le cadre du paradigme de la « vitesse-distance » en design des transports), dans le but d'assurer un pluralisme des enjeux et de rendre visibles des situations encore inconnues ou très émergentes dans un contexte de changement rapide.

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

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

4.4

CAFÉS CITOYENS EN MARGE DES BIBLIOTHÈQUES

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

4.5

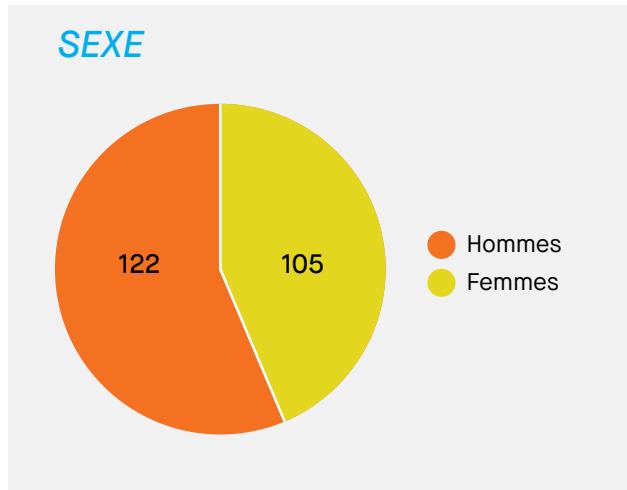
PORTRAIT DES PARTICIPANTS

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

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

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

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



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

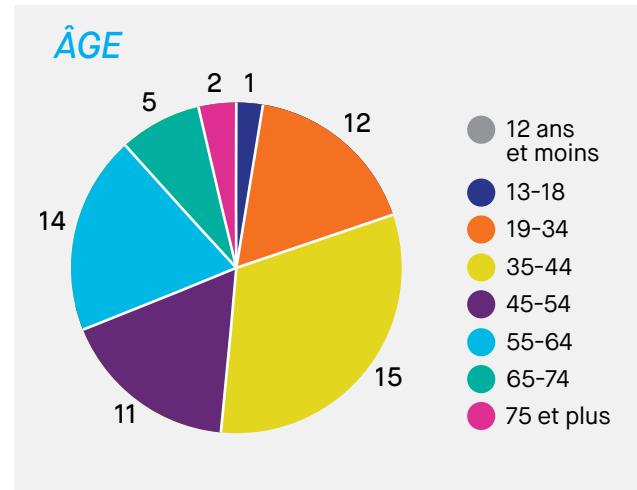


Figure 7 : Répartition des répondants aux cafés citoyens et aux journées de coconstruction par niveau de scolarité atteint

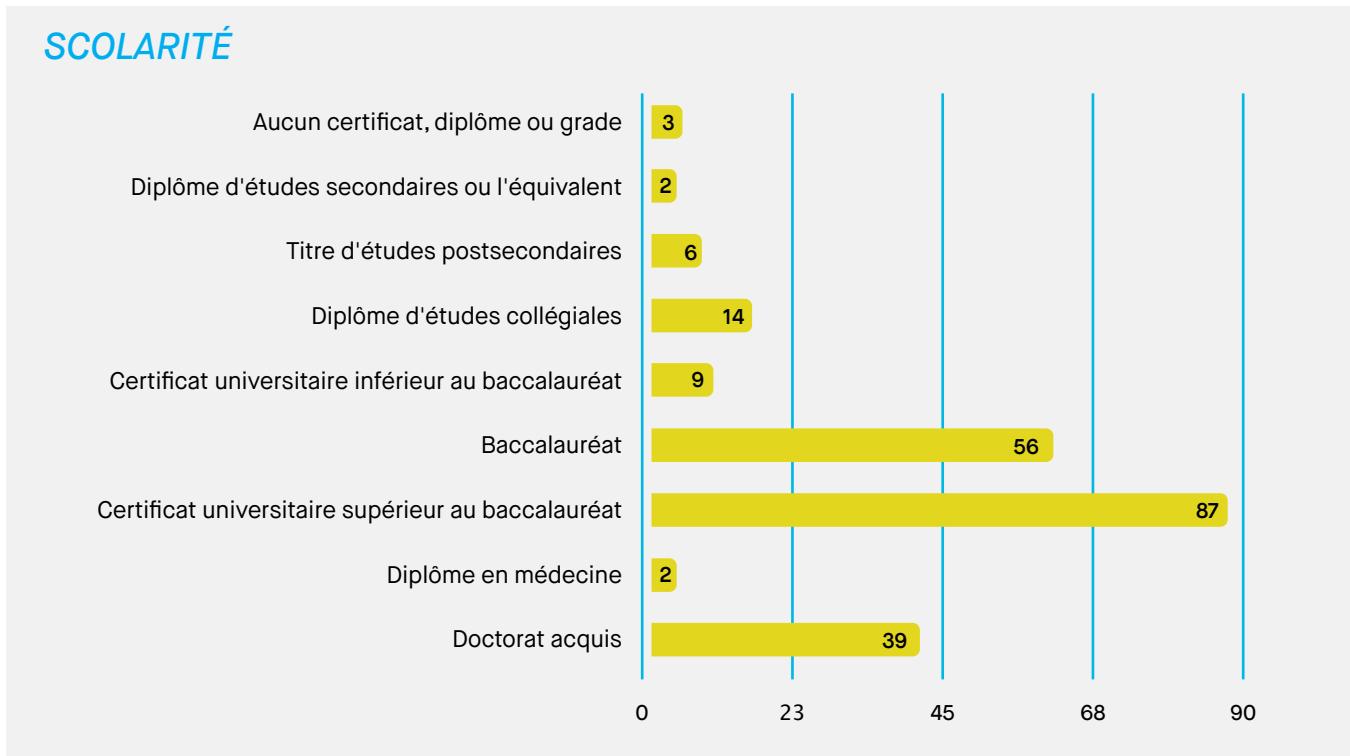
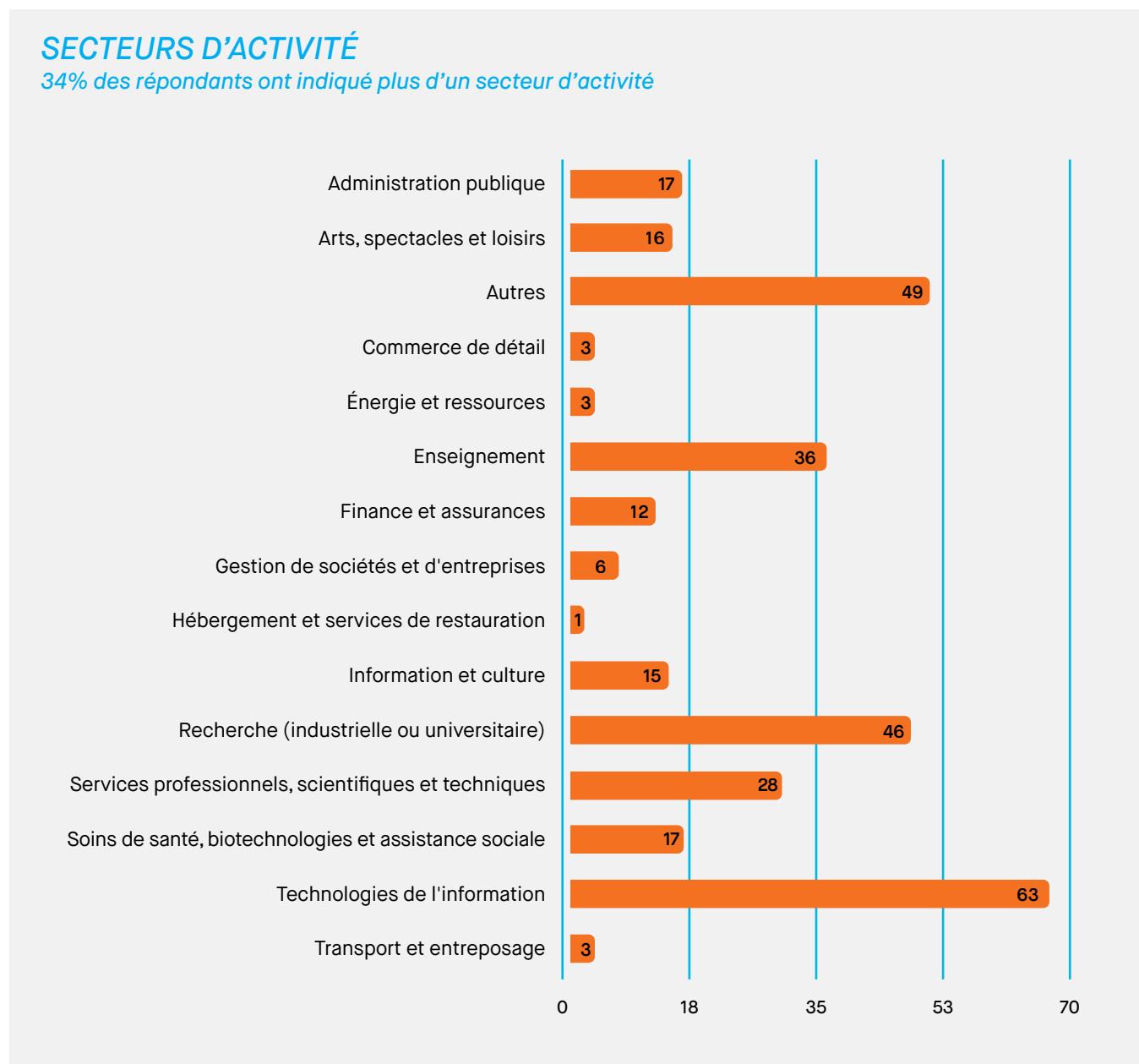


Figure 8 : Répartition des répondants aux cafés citoyens et aux journées de coconstruction par secteur d'activité



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

5.1

LES PARCOURS DÉLIBÉRATIFS

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

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

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

Pour formuler ses propositions, chaque équipe a parcouru trois

étapes où se sont succédées la génération d'idées puis la délibération sur ces idées :

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

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

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

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

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

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

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

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

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

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

Premier moment délibératif : FORMULATION D'ENJEUX ÉTHIQUES EN 2025

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

LE PRINCIPE D'AUTONOMIE

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

LE PRINCIPE DE BIEN-ÊTRE

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

LE PRINCIPE DE DÉMOCRATIE ET DE JUSTICE

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

LE PRINCIPE DE VIE PRIVÉE

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

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

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

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

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

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

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

Pour répondre à ces enjeux, l'équipe a poursuivi ses discussions en essayant de réfléchir ensemble aux quatre principes associés. Plusieurs propositions d'encadrement de l'IA ont été formulées par les

participants. Nous en présentons ici trois (sur six), qui permettent de suivre le cheminement des idées jusqu'à la formulation de la « Une » du Journal.

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

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

Ces propositions, qui dénotent une véritable créativité institutionnelle (au-delà des exemples d'instruments très généraux donnés dans le livret du participant) se situent dans la lignée des enjeux identifiés à l'étape précédente, mais présentent aussi un enrichissement des idées (ce ne sont pas de simples déductions d'instruments adaptés à partir

d'un cas éthique identifié). L'idée de formation à la vigilance et celle de la participation à la décision collective (par un comité multipartite et une planification ouverte) conduisent en effet à des propositions de renforcement des capacités et des formes de démocratie locale.

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

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

« PREMIER ATELIER DE LITTÉRATIE EN MOBILITÉ AUTONOME »

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

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

Le résultat de cette table est particulièrement intéressant car il permet d'envisager la question éthique des véhicules autonome sous l'angle de l'autonomie et de la justice sociale dans la ville, et non exclusivement selon la problématique de la responsabilité dans un scénario d'accident, comme le propose par exemple l'initiative *Moral Machine* du MIT à partir du dilemme moral du tramway²⁹.

²⁹ Site du MIT : <http://moralmachine.mit.edu>

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

Résumé du scénario 2025 de départ.

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

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

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

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

Premier moment délibératif : FORMULATION D'ENJEUX ÉTHIQUES EN 2025

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

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

LE PRINCIPE DE BIEN-ÊTRE

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

LE PRINCIPE D'AUTONOMIE

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

LE PRINCIPE DE RESPONSABILITÉ

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

LE PRINCIPE DE CONNAISSANCE

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

LE PRINCIPE DE JUSTICE

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

LE PRINCIPE DE DÉMOCRATIE

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

Nous pouvons souligner ici que la typologie des principes de la *Déclaration de Montréal* sur l'IA responsable a bien fonctionné pour donner des balises à la discussion, et que les participants ont même proposé des problématiques originales sur certains principes : la nécessité d'aborder le **bien-être** et la **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 |
|---------------------------|---|---|--|
| Description | <ul style="list-style-type: none"> Trop de concentration de pouvoir (cf. GAFAM) ne permettant pas : <ul style="list-style-type: none"> - Le partage équitable des bénéfices de l'IA - L'entrée de nouveaux joueurs (nouveaux modèles d'affaire de type COOP) - La diminution des inégalités (littératie) | <ul style="list-style-type: none"> 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. | <ul style="list-style-type: none"> Définir le bien commun et le type de responsabilité collective dans la transition numérique Par exemple : quelles parties prenantes ? L'entreprise seule ? L'État ? Les syndicats ? Le système éducatif ? |
| Principes associés | Justice et indépendance | Démocratie, connaissance, et autonomie collective | Bien-être et responsabilité |

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

Pour répondre à ces enjeux, l'équipe a poursuivi ses discussions en après-midi avec un nouveau tour de table menant à la rédaction par les participants

de propositions d'encadrement de l'IA sur des *Post-it*, ce qui a conduit à de nombreuses propositions qui ont été discutées une par une collectivement. Le tableau ci-dessous en présente un extrait (six propositions sur plus de dix formulées par le groupe), afin de suivre le cheminement des idées jusqu'à la formulation de la « Une » du Journal.

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

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

Comme dans le cas précédent sur le véhicule autonome, les propositions dénotent une véritable créativité institutionnelle (au-delà des exemples d'instruments très généraux donnés dans le livret du participant). Elles se situent également dans la lignée des enjeux identifiés à l'étape précédente, mais présentent aussi un enrichissement des idées. Si la littératie numérique est bien un objectif à l'agenda de la politique (ex. Stratégie numérique

du Québec), c'est la nécessité qu'elle prenne de l'ampleur qui a été soulignée. Les autres mesures proposées sont inédites et invitent à concevoir de nouveaux dispositifs publics, multipartites ou collectifs pour assurer une véritable autonomie de la société québécoise face aux enjeux de l'IA dans le monde du travail. Dans ce sens, le groupe a pris le parti d'une responsabilité collective face à l'IA dans sa transition vers la société.

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

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

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

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

1. Un nouveau fonds d'assurance numérique doté de 2 milliards (financé par les gains de productivité imputables à l'IA).
2. Une convention avec tous les cégeps et universités pour accélérer les 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. PARTICIPANTS À LA COCONSTRUCTION ET ÉQUIPES DE TRAVAIL

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

| | | | |
|----------------------------|----------------------|------------------------------|---------------------------|
| Sihem Neila Abtroot | Emmanuel Bloch | Jacques Coulombe | Mathieu Dumouchel |
| Sébastien Adam | Marise Bonenfant | Lise Couturier | Benoit Dupont |
| Béatrice Alain | Serge Bouchard | Alexis Cuglietta | Nicolas Dupras |
| Hassane Alami | Caroline Boudreault | Christian Cyr | Diane Duquette |
| Rana Alvabi | Lyne Bourbonnais | Yvonne Da Silveira | Irina Entin |
| Alejandro Arreola-Alvarado | Véronique Boutier | Geneviève Dagneau | Julian Falardeau |
| Gabriel Arruda | Morgane Bravo | Hélène David | Jacqueline Forien |
| Jean-Claude Asssaker | Robert Bruno | François-Michel De Rainville | Simon Frappier |
| Barthélémy Aucourt | Beatrice Cassar | David Décary-Hétu | Benoit Gagnon |
| Naomi Ayotte | Ofelia Castaneda | Guillaume Déraps | Marie-Pierre Gagnon |
| Manon Babine | Chantal Caux | Yves B. Desfossés | Marina Gallet |
| Maryluisa Barillas | Christian Chabot | Michel Desy | Hortense Gallois |
| Philippe Beauchemin | Michel Chabot | Marc-Antoine Dilhac | Sébastien Gambs |
| Stéphane Beaulieu | Karine Charbonneau | Maxime Duban | Véronique Gareau-Chiasson |
| François Beauregard | François Charbonnier | Jean-Yves Dubé | Mathieu Gauthier-Pilote |
| Claude Bédard | Anne Chartier | Geneviève Dubois-Flynn | Sylvie Gélinas |
| Sylvain Bédard | Philippe Chartier | Mathieu Dubreuil-Cousineau | Thomas George |
| Abdelkader Bekhti | Guillaume Chicoisne | Geneviève Dufour | Gueno Gianni |
| Halim Benzaïd | Pierre Choffet | Arnaud Duhoux | Jean-François Gignac |
| Vincent Bergeron | Dominic Cliche | Annie Dulude | Martin Gibert |
| Alexandre Berkesse | Lilen Colombino | Laurence Dumont | Patricia Gingras |
| Karl Bherer | Cristina Cotargasanu | | Béatrice Godard |
| | François Côté | | |

| | | | |
|----------------------|----------------------------|--------------------------|---------------------------|
| Christian Goudreau | Pascale Lehoux | Catherine Olivier | Sara Russo-Garrido |
| Gilles Gouin | Claude Lejeune | Daniel Pascot | Laurence Sabourin |
| Mervine Gowry | Mélanie Levasseur | Florence Paulhiac | Iger Sadoune |
| Alexandre Gravel | Elisabeth Limoges | Ludovic Penet | Marie-Noëlle Saint-Pierre |
| Michel Grou | Pamela Lirio | Jorge Perez | James Sangster |
| Alexandre Guédon | Robert Locas | Caroline Pernelle | Sylvie Saucier |
| Pascaline Guenou | Santiago Lopez | Lorenzo Perozzi | Anton Selikhov |
| Pierre Guillou | Aurélie Macé | Geneviève Perreault | Jean-François Sénéchal |
| François Guité | Aicha Mafhoum | Benoit Petit | |
| Carl Hamilton | Suzanne Mainville | Emmanuel Picavet | Eric Shannon |
| Simon-Pierre Harvey | Mantas Manovas | Louis Piette | Danielle Sicotte |
| Lucie Hébert | Mathieu Marcotte | Frédéric Plamondon | Chantale Simard |
| Ghiles Hellí | Jean-Pierre Marquis | Pier-Luc Plante | Julie Simard |
| Lucas Hubert | Cloderic Mars | Kamila Podgorska-Gilbert | Jean-Hébert Smith-Lacroix |
| Aida Issa | Marie Martel | Keith Poitras | Karima Smouk |
| Sabrina Jocelyn | Mariève Mauger-Lavigne | Julie Politi | Yanis Taleb |
| Erwan Jonchères | Moussa Mekhnach | Philippe Polveche | Isabelle Tanba |
| Nico Julien | Natacha Mercure | Thomas Poulin | Christian Tanguay |
| Debbie Jussome | Bruno Milia | Denis Poussart | Marc Tomkinson |
| Ed Khazen | Michael David Miller | Emmanuelle Praine | Daniel Tremblay |
| Amy Khoury | Ann Mitchell | Louis-Philippe Pratte | Jérémie Trudel |
| Frederic Kleindienst | Erica Monteferrante | Mariel Ramos | Marie-Christianne Trudel |
| Andrée Labrie | Farida Mostefaoui | Diane Raymond | Félix Vaillancourt |
| Anne-Marie Lacombe | Maria Moudfir | Catherine Régis | Julie Verdy |
| Marie-Claude Lagacé | Jocelyne Mouton | Laurence Renault | Arnaud Vicari |
| Henri Lajeunesse | Khalil Mouzawak | Cassie Rhéaume | Danael Villeneuve |
| Karine Landry | Vanessa Murray | Toussaint Riendeau | Grant Wark |
| Jean-Michel Lapointe | Orly Nahmias | Anne-Marie Robert | Bryn Williams-Jones |
| Jonathan Lasprilla | Vanessa Nantel | François-Xavier Robert | Lemy Wong |
| Sylvie Lavoie | John Newhouse | Louis-Nicolas Robert | William Wong |
| Jean Latière | Justin Ngoza | Nicolas Roby | Almina Yagoubi |
| Louis Lecaer | Zoonie Nguyen | Stéphane Roche | Ming Yue |
| Dominique Leclerc | Lisa Marlène Ntibayindusha | Marie Roy | |
| Sarah Legendre | | | |
| Bilodeau | | | |

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

Simon Beaudoin-Gagnon, Maison des étudiants canadiens

Alexandre Beaudoin-Peña, Université de Montréal

Bhavish Beejan, Université Laval

Liam Bekirsky, Maison des étudiants canadiens

Karl Bherer, Université Laval

Alexis Bibeau, Université Laval

Pierre-Antoine Boutin-Panneton, Université Laval

Katie Charpentier-Bourque, Maison des étudiants canadiens

Arnaud Brubacher-Chouinard, Maison des étudiants canadiens

Dominic Cliche, Université Laval

Valentine Crosset, Université de Montréal

Rosemarie Desmarais, Maison des étudiants canadiens

Eve Gaumond, Université Laval

Martin Gibert, IVADO et Centre de recherche en éthique éthicien

Emilie Guiraud, Université Laval

Haykuhi Gutrez, Maison des étudiants canadien

Hubert Hamel-Lapointe, Université de Montréal

Audrey Houle, Université Laval

Samira Illourman, Maison des étudiants canadiens

Nico Julien, Université Laval

Henri Lajeunesse, Université Laval

Lauriane Long-Raymond, Maison des étudiants canadiens

Guillaume Macaux, Université Laval

Vincent Mai, Université de Montréal

Mariève Mauger-Lavigne, Université de Montréal

Christophe Mondin, CIRANO

Orly Nahmias, citoyenne

Pauline Noiseau, Université de Montréal

Judith Paquet, Université Laval

Pierre-Luc Plante, Université Laval

Léa Ricard, Université de Montréal

Lynda Robitaille, Centre de recherche en données massives, Université Laval

Jason Stanley, Université de Montréal

Yanis Taleb, Université de Montréal

Clémence Varin, Université Laval

Nathalie Voarino, Université de Montréal

Camille Vézy, Université de Montréal

Alessia Zarzani, Université de Montréal

Les experts consultés

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

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

Guillaume Chicoisne, directeur des programmes scientifiques, IVADO

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

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

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

Jihane Lamouri, coordonnatrice à la diversité, IVADO

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

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

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

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

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

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

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

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

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

L'équipe de gestion

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

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

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

L'équipe de recherche et analyse

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

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

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

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

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

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

Christophe Mondin, professionnel de recherche chez CIRANO

Camille Vézy, doctorante en communication, Université de Montréal

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

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

L'équipe de coordination à Paris

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

Hanane Hadjiloum, chargée des communication Maison des étudiants canadiens

Christine Métayer, directrice de la Maison des étudiants canadiens

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

Les partenaires ayant contribué à la coconstruction de l'automne

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

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

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

ANNEXES

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

Les cafés citoyens

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

Le café citoyen (« world café ») est un dispositif de conversation convivial visant à faciliter le dialogue constructif et le partage d'idées. On recherche l'ambiance d'un café où les participants débattent d'une question en petits groupes. À intervalles réguliers, les participants changent de table. Un hôte reste à la table et résume la conversation précédente aux nouveaux arrivés. Les conversations en cours

sont alors « fécondées » par les idées issues de la conversation précédente. Au terme du processus, les principales idées sont résumées lors d'une assemblée plénière, et les possibilités de suivi sont soumises à la discussion³⁰.

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

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

Ci-dessous, le déroulement type des cafés citoyens :

Tableau 5 : Déroulement type des cafés citoyens

| Étapes | Heure | Description |
|---|----------------|---|
| Accueil | 13 h à 13 h 30 | Cafés et collations |
| Découverte de l'IA et de ses enjeux éthiques et sociaux | 13 h 30 à 14 h | Introduction pédagogique Introduction aux enjeux éthiques et sociaux de l'intelligence artificielle (<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. |

³⁰ Définition provenant de l'Institut du nouveau monde

Les journées de coconstruction

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

L'approche prospective retenue pour les présents travaux consiste à se projeter dans un futur proche

(2025) pour opérer un détour imaginaire et penser ensuite rétrospectivement des chemins innovants pour relier le présent aux futurs les plus souhaitables.

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

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

Tableau 6 : Déroulement type des journées de coconstruction

| Étapes | Heure | Description |
|--|-------------------|---|
| Accueil | 8 h 30 à 9 h | Café et viennoiseries |
| Mot de bienvenue et découverte de l'IA | 9 h à 10 h | Introduction : principes de l'intelligence artificielle, les enjeux éthiques de l'IA (<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 |

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

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

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

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

ÉQUIPE DE RÉDACTION DES SCÉNARIOS

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

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

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

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

Vincent Mai, doctorant en robotique, Université de Montréal

Christophe Mondin, professionnel de recherche chez CIRANO

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

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

1. L'ensemble des scénarios résumés par thème

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

Tableau 7 : Résumé des scénarios

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

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

2. Cinq scénarios complets

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

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

Tableau 8 : Constitution de cinq scénarios par thème

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

Thème 1 : SANTÉ PRÉDICTIVE

Scénario de départ : LES JUMEAUX NUMÉRIQUES

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

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

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

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

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

Thème 2 : VILLE INTELLIGENTE

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

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

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

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

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

Thème 3 : ÉDUCATION PRÉDICTIVE

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

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

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

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

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

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

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

Thème 4 : JUSTICE ET POLICE PRÉdictIVE

Scénario variante : DÉCISION DE LIBÉRATION CONDITIONNELLE

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

Connaissant l'allergie sévère de Jean aux arachides, son frigo, qui communiquait sa liste d'épicerie à un magasin partenaire, la formatait en fonction de cette caractéristique. Toutefois, lorsque Sylvia a piraté le système, l'allergie aux arachides de Jean n'apparaissait plus dans les paramètres par défaut et le frigo a produit une liste inadaptée à ses besoins

de santé. En mangeant un plat préparé contenant une faible dose d'arachides, Jean a commencé à avoir des difficultés à respirer et a dû se rendre d'urgence à l'hôpital.

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

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

- > Des facteurs historiques statiques, à savoir l'âge auquel Sylvia a commis sa première infraction et ses antécédents criminels (Sylvia avait déjà piraté la commande du pilulier de sa mère à 18 ans, et le réseau de caméras de vidéosurveillance de son quartier à 25 ans) ;
- > Des facteurs de risque dynamiques : l'occupation de Sylvia, ses fréquentations, ses relations amoureuses et familiales, les remords exprimés par Sylvia, etc.

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

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

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

Thème 5 : MONDE DU TRAVAIL

Scénario de départ : UNE RESTRUCTURATION SOCIALEMENT RESPONSABLE

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

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

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

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

CRÉDITS DU RAPPORT FINAL

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

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

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

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

Coordination

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

Collaboration aux contenus

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

Révision et édition

Chantal Berthiaume, gestionnaire de contenu et rédactrice

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

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

Traduction

Rachel Anne Normand et François Girard, Services linguistiques Révidaction

Graphisme

Stéphanie Hauschild, directrice artistique

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

NOS PARTENAIRES



CENTRE DE RECHERCHE EN ETHIQUE



ICRA
Programme
IA et
société



Centre interuniversitaire de recherche en analyse des organisations



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



Canada



Canadian
Cultural
Centre
Paris



