

Penser l'intelligence artificielle responsable : un guide de délibération



Penser l'intelligence artificielle responsable : un guide de délibération

Introduction	2
<hr/>	
Première partie	
Intelligence artificielle	4
Qu'est-ce que l'IA?	4
Ce que fait l'IA	5
Ce que ne fait pas l'IA	10
<hr/>	
Deuxième partie	
L'éthique de l'IA	11
Qu'est-ce que l'éthique de l'IA ?	11
Quelques enjeux éthiques et sociaux	12
Des principes à l'action	14
Déclaration de Montréal IA responsable	15
<hr/>	
Troisième partie	
Délibérer sur l'éthique de l'IA	17
Impliquer les citoyennes et les citoyens	17
Pourquoi délibérer sur l'éthique de l'IA ?	18
<hr/>	
Crédits	21
Remerciements	22

Algorithmes, données, intelligence artificielle (IA) : autant de termes qui font désormais partie de notre quotidien, mais qui demandent un peu d'explications. Comment comprendre ces technologies ? Certes, elles suscitent un engouement très large et sont porteuses de promesses pour un avenir meilleur. Mais quels sont les risques encourus si on perd la maîtrise de leur développement ?

Pour relever les défis d'un développement responsable de l'IA, de nombreux organismes privés, institutions publiques et organisations internationales ont publié des chartes de bonnes pratiques, des déclarations de principes et des recommandations. Ils ont permis de constater des convergences autour de principes clés (justice, autonomie, bien-être). Cependant, les principes sont parfois abstraits et ne sont pas toujours définis de la même façon à travers le monde.

Il reste donc beaucoup à faire. Il s'agit à présent de réfléchir à la mise en œuvre des principes éthiques généraux, de les rendre applicables à des secteurs d'activités spécifiques (éducation, science, information, santé, etc.) et de mettre concrètement en pratique les propositions issues de cette réflexion. Ensuite, il est indispensable d'impliquer davantage les citoyennes et les citoyens pour définir les lignes directrices d'un usage responsable de l'IA et des données massives et, de recueillir leurs opinions informées. En effet, le déploiement de l'IA affecte l'ensemble des citoyennes et des citoyens et pose des questions éthiques et politiques qui devraient faire l'objet d'une délibération citoyenne. Enfin, il est primordial de consolider la littératie numérique du grand public, c'est-à-dire d'informer et d'outiller les citoyennes et les citoyens pour leur permettre de mieux comprendre les enjeux

du développement responsable de l'IA et de participer aux délibérations publiques sur les principes et les normes de son déploiement.

La participation aux ateliers de délibération sur l'IA et les technologies numériques repose sur une bonne compréhension des enjeux éthiques et sociétaux de l'IA et des règles de la délibération. Le présent guide a précisément pour vocation de rendre accessible l'IA et les questions éthiques qui s'y rapportent, et d'introduire à la délibération sur l'éthique de l'IA. Ce guide comprend des définitions, des illustrations et des cas d'usage. Il permet ainsi de créer un langage commun autour de l'éthique de l'IA. Ce guide a aussi pour ambition d'outiller les communautés pour qu'elles organisent leurs propres délibérations sur l'IA et le déploiement des technologies numériques dans leur environnement social. Il est

conçu pour permettre à des enseignantes et des enseignants, à des représentantes et des représentants communautaires, à des citoyennes et des citoyens, et à des administratrices et des administrateurs de faire vivre le débat et de définir des stratégies communes.

Cette démarche délibérative et participative repose sur la confiance dans les capacités des individus à concevoir leur avenir et le type de société dans laquelle ils et elles souhaitent vivre, à formuler les principes éthiques et politiques qui devraient l'organiser et à élaborer des propositions pertinentes de politique publique.

Enfin, ce document présente certaines limites inévitables. Il se veut simple pour plus de clarté et d'efficacité, mais il est aussi teinté culturellement. C'est pourquoi il sera adapté aux différentes réalités géographiques et culturelles dans lesquelles seront organisées des délibérations. Tout le monde est invité à l'enrichir. Nous formons l'espoir que ce guide favorise la délibération entre citoyennes, citoyens, parties prenantes et personnes en charge des affaires publiques, et que les ateliers et forums de délibération qu'il facilite contribuent à un développement plus responsable et plus démocratique de l'IA.

L'équipe d'Algora Lab



Première partie

Intelligence artificielle

QU'EST-CE QUE L'IA?

L'IA est l'ensemble des techniques informatiques qui permettent à une machine (p. ex. un ordinateur, un téléphone) de réaliser des tâches qui requièrent généralement de l'intelligence, comme le raisonnement ou l'apprentissage. On parle aussi d'automatisation de tâches intelligentes. Les avancées scientifiques dans le domaine de l'IA, comme les techniques d'apprentissage profond, ont

permis de mettre au point des machines intelligentes très performantes grâce à l'accès à d'énormes quantités de données et à une puissance de calcul de plus en plus élevée. Ces nouvelles techniques ont été rapidement déployées à grande échelle, dans tous les domaines de la vie sociale, dans le transport, l'éducation, la culture et la santé.



CE QUE FAIT L'IA

L'intelligence artificielle repose sur l'utilisation d'algorithmes qui traitent des données. Un algorithme est une suite d'instructions qui permettent de résoudre des problèmes et d'accomplir des tâches complexes. Cette série d'étapes permet de transformer une information d'entrée (input) en un résultat utile (output). Une recette de cuisine est une

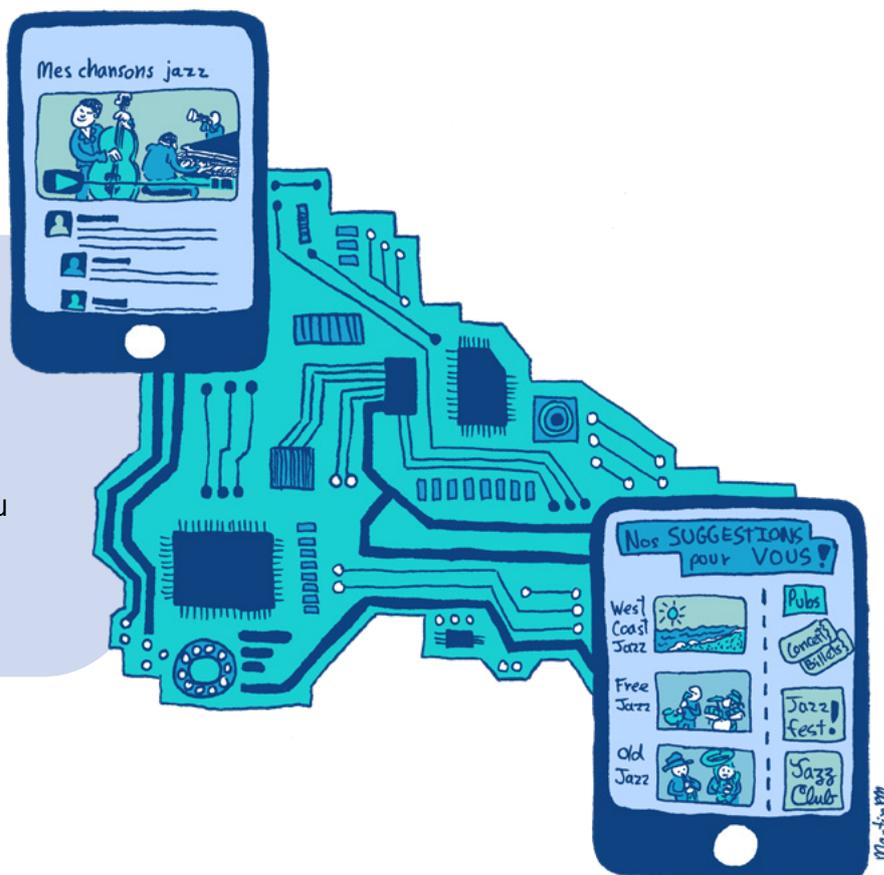
sorte d'algorithme : pour cuisiner un plat (output), il faut avoir les bons ingrédients (input) et suivre une recette qui donne des instructions pour les utiliser correctement (l'algorithme). La suite d'instructions qu'un ordinateur utilise pour prédire l'âge d'une personne (output) à partir de sa photo (input) est aussi définie par un algorithme.

Les algorithmes développés dans le domaine de l'IA permettent de réaliser des fonctions.

En voici quelques exemples.

LA PRÉDICTION

(p. ex., prédire l'intérêt d'une utilisatrice ou d'un utilisateur d'internet pour un type de contenu culturel à partir de ses données de navigation)



LA DÉTECTION

(p. ex. détecter s'il y a un visage et à quel endroit à partir d'une photo)

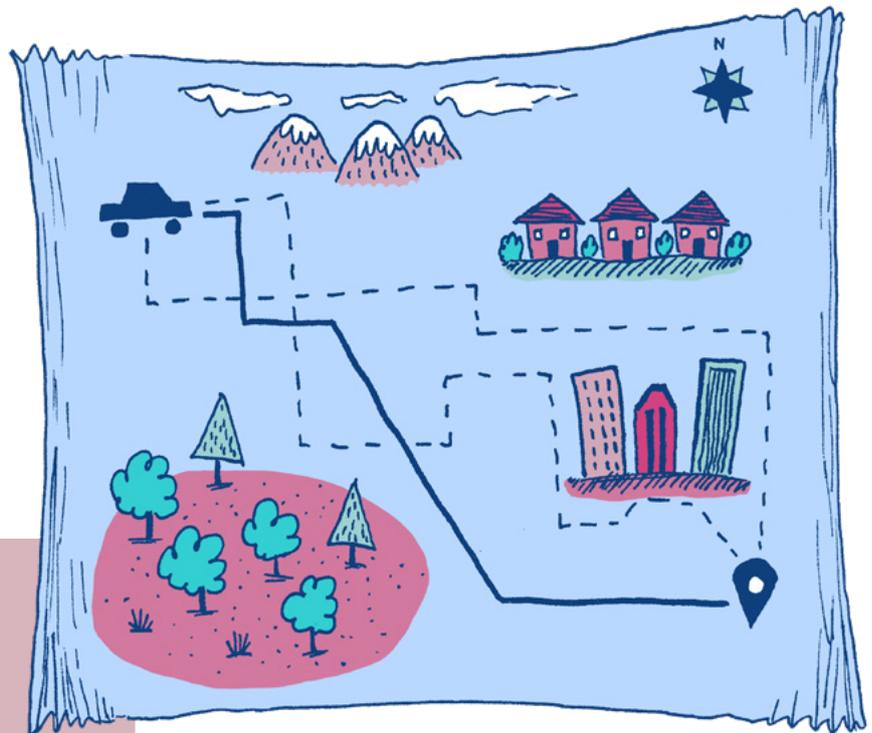
L'IDENTIFICATION

(p. ex., trouver le nom d'une personne à partir de sa photo)



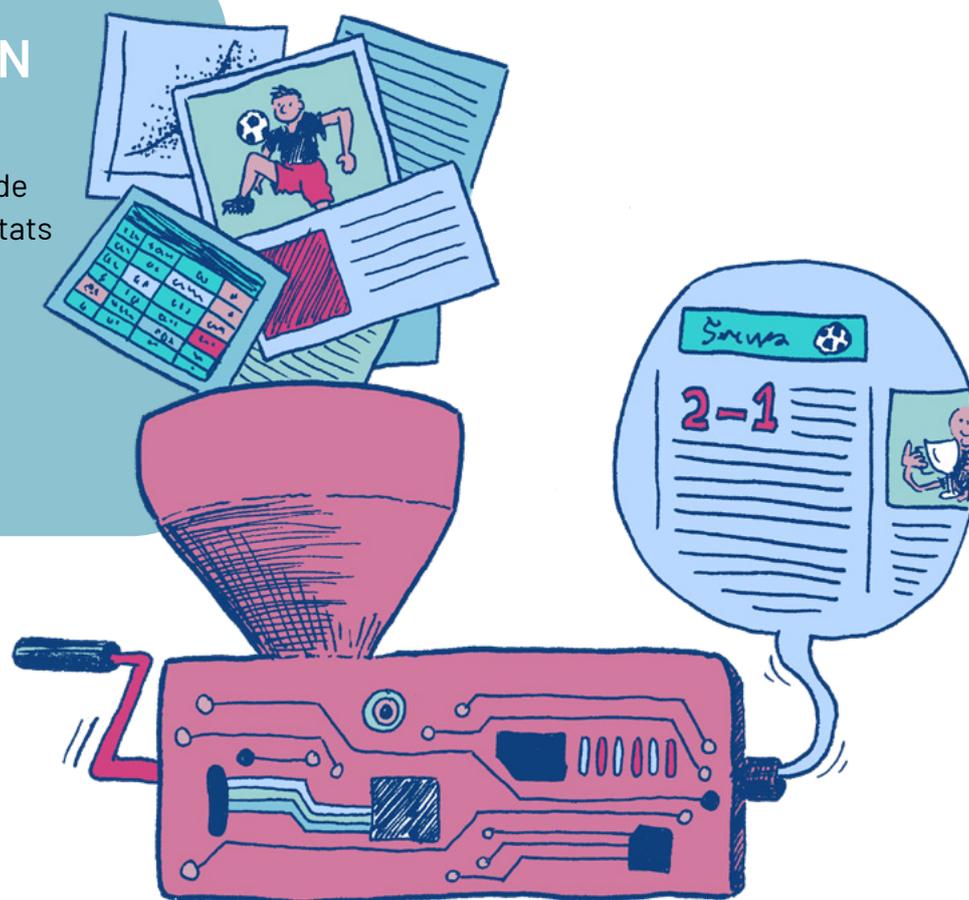
LA PLANIFICATION ET LA PRISE DE DÉCISION

(p. ex., choisir le trajet le plus rapide pour aller à l'hôpital à partir de données de circulation)



LA GÉNÉRATION DE CONTENU

(p. ex., générer un article de journal à propos des résultats d'un événement sportif)



La combinaison de ces fonctions permet d'accomplir plusieurs tâches de haut niveau.

En voici quelques exemples.

> LA PERCEPTION

Lorsque l'IA est utilisée pour analyser les mesures, comme l'image d'une caméra ou l'enregistrement d'un microphone. Un algorithme peut ainsi reconnaître des mots dits à voix haute ou écrits à la main, ou encore déterminer s'il y a un rassemblement de personnes à partir d'une photographie.

> LE TRAITEMENT DE LANGAGE NATUREL

Lorsque l'IA est utilisée pour traiter la signification des mots. Une IA peut comprendre un ordre qui lui est donné, comme appeler un contact ou prédire le prochain mot qui sera écrit, comme sur les claviers des téléphones intelligents. Il est aussi possible de détecter si un article de nouvelles intéressera une personne donnée, ou encore de générer la traduction d'un paragraphe dans une langue étrangère.

> LA ROBOTIQUE

Lorsque l'IA contrôle une machine qui peut agir dans le monde physique. À partir de l'information reçue par ses capteurs, l'IA doit planifier et prendre la meilleure décision afin d'accomplir son objectif. Par exemple, une voiture autonome doit conduire de manière sécuritaire jusqu'à une destination donnée, et un robot chirurgical doit répliquer au mieux les mouvements de l'humain qui le contrôle.

> L'OPTIMISATION ET LA RÉOLUTION DE PROBLÈME

Lorsque l'IA doit résoudre une situation dans un environnement défini. Par exemple, prendre des décisions dans un jeu vidéo ou planifier le trajet optimal pour une voiture ou les déplacements de taxis.

Une grande variété de **domaines d'application** peuvent bénéficier de la possibilité d'automatiser ces fonctions : l'éducation, le journalisme, la cybersécurité, les jeux vidéo, l'art (musique, cinéma...), la finance, les soins de santé, les transports, le militaire, l'écologie, les sciences du climat, etc.

POUR ALLER PLUS LOIN

Il y a **deux branches techniques** principales en IA. Les conceptrices et les concepteurs de l'algorithme peuvent soit donner toutes les **instructions** à la machine à l'avance, ou alors permettre à la machine d'**apprendre** les étapes par elle-même. Écrire les règles à l'avance requiert des conceptrices et des concepteurs qu'ils sachent eux-mêmes **comment résoudre** le problème qu'ils demandent à l'IA de résoudre. Dans certains cas, c'est possible : c'est par exemple grâce à des modèles scientifiques basés sur les lois de la physique que l'on prédit la météo. Mais lorsque le problème est trop **complexe**, les conceptrices et les concepteurs de l'IA préfèrent utiliser l'apprentissage machine. Dans ce cas, le modèle doit être **entraîné**. Par exemple, il est très difficile d'écrire les règles qui permettent à un ordinateur de différencier un chat d'un chien sur une image, puisqu'il y a une grande variation dans ces deux espèces, et en plus, les photos peuvent être prises d'angles différents, sous différents éclairages, etc. Les algorithmes d'apprentissage machine permettent de résoudre le problème en montrant plusieurs images de chiens et de chats à un ordinateur, et en lui apprenant à les différencier par lui-même. Bien sûr, la performance de l'algorithme dépendra de la **quantité** d'images ainsi que de leur **représentativité** – si on l'a entraîné avec seulement des chats noirs vus de face, l'algorithme ne saura que faire d'une photo d'un chat blanc prise du dessus. C'est pourquoi les algorithmes d'IA ont généralement besoin d'une grande quantité de données et que la manière dont elles auront été collectées importe.

CE QUE NE FAIT PAS L'IA

À l'heure actuelle, il n'existe pas d'IA générale, ou « forte », qui, à l'image de l'intelligence humaine, serait capable d'accomplir des tâches aussi différentes que jouer aux échecs, conduire une voiture et reconnaître des tumeurs. C'est un objectif pour certaines chercheuses et certains chercheurs du domaine, mais les systèmes actuels les plus avancés en sont encore loin. L'IA d'aujourd'hui est qualifiée de « faible », car si elle peut réaliser certaines tâches de manière plus performante qu'un être humain, ce n'est que sur les tâches précises pour lesquelles elle a été conçue. Certains envisagent aussi le fait qu'une IA générale pourrait un jour avoir des émotions ou une conscience d'elle-même. Nous n'en sommes pas là, et tout ce que peut faire une IA faible consiste à identifier des émotions et à les simuler.

NOTIONS ESSENTIELLES

INTERNET DES OBJETS

Infrastructure d'objets interconnectés et capables de communiquer entre eux sans intervention humaine.

ALGORITHME

Suite d'instructions qui permet de transformer une entrée en une sortie. Par exemple, une recette de crêpes permet de transformer des ingrédients en un repas délicieux en suivant des étapes précises. Les étapes pour résoudre un cube Rubik sont aussi un algorithme.

DONNÉES MASSIVES, MÉGADONNÉES OU BIG DATA

Ensemble de données si volumineuses qu'elles ne peuvent pas être collectées, stockées et analysées avec les méthodes traditionnelles. Beaucoup d'algorithmes d'IA utilisent des mégadonnées.

APPRENTISSAGE MACHINE

Capacité d'une machine d'apprendre à réaliser une tâche sans instructions, mais plutôt grâce à une expérience acquise durant un procédé d'entraînement.

APPRENTISSAGE PROFOND

Apprentissage machine avec une technique spécifique et particulièrement puissante qui utilise des réseaux de neurones connectés sur plusieurs couches. Ces réseaux de neurones artificiels, inspirés du fonctionnement des neurones du cerveau, consistent en un traitement mathématique des données entrantes.

Deuxième partie

L'éthique de l'IA

Si les technologies d'IA se montrent prometteuses, leur utilisation présente des enjeux éthiques et sociaux sur lesquels il est essentiel de réfléchir collectivement, considérant leur influence croissante sur nos sociétés. C'est un des objectifs de l'éthique appliquée à l'IA, qui cherche à cerner et à empêcher les mauvaises utilisations de l'IA tout en favorisant les bénéfices qui en découlent.

QU'EST-CE QUE L'ÉTHIQUE DE L'IA ?

L'éthique désigne un ensemble de normes (principes, règles) auxquelles nous devons nous soumettre si nous voulons bien agir, comme le principe selon lequel on ne doit pas causer de tort à autrui ou la règle qui interdit le mensonge. On dit que l'éthique est prescriptive parce qu'elle prescrit ce qui doit être, ce qu'il faut faire ou encore ce qui est acceptable selon les valeurs adoptées. Les normes éthiques articulent des valeurs que nous reconnaissons comme morales. L'éthique désigne aussi la discipline philosophique qui tente d'établir ces normes et ces valeurs.

L'éthique de l'IA est l'ensemble des normes et des valeurs appliquées au développement et à l'utilisation de l'IA. Il s'agit donc d'un domaine restreint d'application de l'éthique, mais comme les technologies de l'IA bouleversent l'organisation sociale et peuvent avoir des conséquences nuisibles très profondes, ce domaine de l'éthique apparaît comme crucial et connaît un développement très important.

Enfin, l'éthique de l'IA fait partie de l'éthique publique, c'est-à-dire l'éthique appliquée à des controverses sociétales qui requièrent une solution acceptable pour les individus qui sont en désaccord. C'est le cas aussi de l'éthique de l'environnement ou de la bioéthique.

QUELQUES ENJEUX ÉTHIQUES ET SOCIAUX

Expertes et experts, chercheuses et chercheurs, ainsi que citoyennes et citoyens ont déjà exprimé de nombreuses préoccupations quant au développement et à l'utilisation de l'IA. Parmi celles-ci, on retrouve :

> LE RISQUE D'ATTEINTE À LA VIE PRIVÉE

La vie privée est une valeur sociale apparue récemment dans les sociétés qui aspiraient à la démocratie. Elle est désormais largement acceptée et nombreuses sont les personnes qui s'inquiètent de la violation de leur vie privée. Le risque d'atteinte à la vie privée est favorisé par la performance inédite de moyens de collecte qui sont de plus en plus intrusifs et par la multiplication des lieux où sont collectées ces données personnelles (p. ex., les téléphones portables ou les objets connectés dans les domiciles). Ce risque est également accru en raison des nouvelles possibilités d'analyses qu'offre l'IA. Par exemple, les algorithmes peuvent identifier des individus en croisant différentes données qui avaient été anonymisées.

> LE RISQUE DE LIMITER LA LIBERTÉ DE CHOIX ET L'AUTONOMIE

On valorise généralement la liberté de choix et l'autonomie, c'est-à-dire la capacité à être l'autrice ou l'auteur de ses décisions. Or, les machines peuvent considérablement restreindre notre liberté en prenant des décisions pour nous sans que nous en ayons conscience ou que nous puissions les contester. On peut penser aux effets de « bulle filtrante » des algorithmes qui proposent aux utilisatrices et aux utilisateurs du contenu toujours conforme à leur comportement numérique (c'est-à-dire leurs précédents choix), limitant ainsi la diversité du contenu qui leur est proposé ou les chances de découvrir de nouvelles préférences. C'est le cas lorsque des sites de contenu musical proposent des morceaux toujours similaires à ce que l'utilisatrice ou l'utilisateur écoute.

> LE RISQUE DE DISCRIMINATION

La discrimination consiste à traiter de manière différente des cas semblables sans bonne raison. Elle est le contraire de la justice, qui consiste à traiter également des cas semblables. Une machine intelligente peut reproduire ou engendrer des discriminations quand son algorithme ou les données sur lesquelles elle apprend contiennent des erreurs ou des biais qui conduisent à traiter différemment, sans justification acceptable, des personnes ou des

groupes par rapport au reste de la population. Ainsi, des personnes ou des groupes se trouvent exclus des bénéfices de l'IA. Une machine qui est entraînée sur des données qui ne portent pas sur toute la population va priver une partie de cette population de ses bénéfices; c'est le cas lorsqu'une machine qui détecte les cancers de la peau n'est entraînée que sur des peaux claires : elle ne détectera pas avec la même fiabilité les cancers qui apparaissent sur des peaux plus sombres.

> LE MANQUE D'EXPLICABILITÉ

Il est souvent difficile d'expliquer comment un algorithme est parvenu à telle décision ou telle recommandation. Les machines intelligentes sont comme des « boîtes noires » dont on ne comprend pas le fonctionnement. Or, il paraît important d'être en mesure de comprendre une décision prise par un algorithme pour pouvoir garder le contrôle des décisions qui nous affectent, pour pouvoir aussi la contester, voire la changer. L'explicabilité est une valeur étroitement liée à la liberté de choix et à l'autonomie, mais aussi à la justice, comme lorsque la machine intelligente reproduit ou engendre des discriminations.

> L'AUTOMATISATION DU TRAVAIL

Alors que les machines intelligentes permettent d'effectuer des tâches répétitives et peuvent ainsi réduire la pénibilité du travail, elles tendent

à remplacer les travailleuses et les travailleurs humains. Outre les conséquences sociales et économiques de ce remplacement, comme le chômage grandissant, ce remplacement affecte le sens de la solidarité et du respect de soi. Un des enjeux éthiques est alors le maintien de relations humaines, notamment dans des secteurs comme la santé ou l'éducation; un autre est celui de la préservation des conditions du respect et de l'estime de soi qui reposent en partie sur le sentiment de son utilité sociale.

> LES RISQUES POUR L'ENVIRONNEMENT

La protection de l'environnement et la lutte contre le changement climatique constituent désormais des enjeux sociaux majeurs. Le déploiement de l'IA peut aider à améliorer nos actions collectives et individuelles pour réduire notre impact négatif sur l'environnement, mais il peut également, si on ne s'en préoccupe pas, augmenter cet impact négatif. S'intéresser à l'enjeu environnemental de l'IA implique de prendre en considération l'impact de l'ensemble des objets qui rendent possible l'utilisation de l'IA, comme les téléphones intelligents avec leurs composants polluants ou les centres de données. Par exemple, les centres de stockage des mégadonnées (data centres), qui permettent notamment l'apprentissage machine, sont de grands consommateurs d'énergie.

DES PRINCIPES À L'ACTION

Afin d'apporter des réponses socialement acceptables aux différentes préoccupations éthiques, sociétales et politiques que suscite le déploiement de l'IA, de nombreuses déclarations de principes éthiques et lignes directrices ont été produites à travers le monde. Ces documents tentent de définir les principes éthiques qui permettront de guider la réflexion afin de limiter les conséquences négatives de l'utilisation de l'IA. C'est par exemple le cas des principes adoptés par l'Organisation de coopération et de développement économiques (OCDE) et les pays du G20, des lignes directrices pour une IA digne de confiance de la Commission européenne, ou encore de la Déclaration de Montréal pour un développement responsable de l'IA, qui se distingue par le processus délibératif qui a alimenté le travail des expertes et des experts. Cette déclaration promeut 10 principes éthiques pour un développement responsable de l'IA, comme le principe de bien-être :

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

Les principes ont pour vocation de guider l'action. Ils peuvent ainsi être à la base de différentes mesures concrètes de la gestion et du développement éthique de l'IA. Il peut s'agir, par exemple :

- > De lois;
- > De politiques publiques;
- > D'audits ou de certifications;
- > De formations;
- > D'acteurs institutionnels;
- > De codes de conduite ;
- > De solutions techniques.

NOTIONS ESSENTIELLES

DILEMME ÉTHIQUE

Une situation présente un dilemme éthique quand elle exige un choix entre deux options incompatibles et toutes deux moralement insatisfaisantes. Cela arrive quand il y a un conflit de principes ou de valeurs et que l'on doit alors sacrifier une valeur morale importante.

GOVERNANCE ALGORITHMIQUE

Le terme peut référer soit à la manière de gouverner les algorithmes, d'encadrer leur développement et de réguler leur utilisation, soit à la manière dont les algorithmes nous gouvernent ou régulent notre vie, nos relations sociales, nos institutions publiques.

BOÎTE NOIRE

On désigne par le terme de boîte noire un système d'IA dont il est difficile ou impossible d'expliquer les décisions ou les recommandations. Ce terme souligne le manque de transparence du fonctionnement des machines intelligentes, notamment celles qui utilisent des algorithmes d'apprentissage profond (*deep learning*).

ÉTHIQUE DÈS LA CONCEPTION (*ethics by design*)

Une manière d'encadrer l'usage éthique des outils numériques et des systèmes d'IA consiste à prendre en considération les principes éthiques dès leur conception par les chercheuses et les chercheurs ainsi que les ingénieures et les ingénieurs. On empêche ainsi en amont leurs usages non éthiques ou non désirables socialement.

BIAIS

On entend par biais le mécanisme de la pensée qui altère et fausse un jugement. Un algorithme peut présenter des biais s'il favorise certains résultats sans une bonne justification morale.

LA DÉCLARATION DE MONTRÉAL IA RESPONSABLE

La Déclaration de Montréal IA responsable (la Déclaration) est une liste cohérente de dix principes éthiques qui a pour objectif de donner des repères moraux pour guider la réflexion sur l'intelligence artificielle responsable et socialement acceptable, et d'encadrer son développement, son déploiement et son utilisation dans différents secteurs de la vie sociale.

La Déclaration est issue d'un processus délibératif inclusif rassemblant citoyens, experts, responsables publics, parties prenantes de l'industrie, organisations de la société civile et ordres professionnels. Cette initiative de l'Université de Montréal a rassemblé plus de 500 participants dans des ateliers de délibération.

L'intérêt de cette démarche délibérative est triple :

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

Les principes ne sont pas hiérarchisés. Le dernier principe n'est pas moins important que le premier. 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. Cela permet une diversité légitime d'interprétations mais cela n'autorise pas n'importe quelle interprétation.

Il faut aussi souligner le fait que ce sont des principes éthiques et non des règles de gouvernance, encore moins des normes juridiques. Cependant, ils peuvent être traduits en langage politique et interprétés de manière juridique, notamment dans le langage des droits humains fondamentaux.

Enfin, bien que ces principes aient été élaborés dans une société donnée, la société québécoise et canadienne, ils constituent une base pour un dialogue interculturel et international.

La Déclaration est le document d'éthique de l'IA le plus complet à ce jour et reste ouverte à discussion et à révision.

En voici la liste des dix principes qui se déclinent en une soixantaine de sous-principes (à découvrir sur le site : declarationmontreal-iaresponsable.com)

1. PRINCIPE DE BIEN-ÊTRE

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

2. PRINCIPE DE RESPECT DE L'AUTONOMIE

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

3. PRINCIPE DE PROTECTION DE L'INTIMITÉ ET DE LA VIE PRIVÉE

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

4. PRINCIPE DE SOLIDARITÉ

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

5. PRINCIPE DE PARTICIPATION DÉMOCRATIQUE

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

6. PRINCIPE D'ÉQUITÉ

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

7. PRINCIPE D'INCLUSION DE LA DIVERSITÉ

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

8. PRINCIPE DE PRUDENCE

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

9. PRINCIPE DE RESPONSABILITÉ

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

10. PRINCIPE DE DÉVELOPPEMENT SOUTENABLE

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

Troisième partie

Délibérer sur l'éthique de l'IA

IMPLIQUER LES CITOYENNES ET LES CITOYENS

Il existe plusieurs façons d'inclure les citoyennes et les citoyens dans le débat sur l'éthique de l'IA. Selon leur degré d'implication, on peut généralement distinguer trois processus types :

> CONSULTER

La consultation est un processus qui consiste à recueillir les opinions déjà formées des personnes consultées sur un thème préalablement défini. La consultation permet aux personnes consultées de poser des questions, d'exprimer leurs préoccupations, leurs attentes, leurs commentaires ou leurs opinions dans le but d'améliorer le processus décisionnel.

> DÉLIBÉRER

La délibération est une discussion raisonnée à travers un échange d'arguments en vue d'une décision collective. La délibération doit permettre d'accroître la connaissance de chacune et de chacun, de comprendre ses propres intérêts et les intérêts du groupe. Elle peut conduire à la modification de nos préférences initiales. Elle ne vise pas nécessairement le consensus, mais l'identification d'orientations communes fondées sur les convergences et les divergences d'opinions et les raisons qui les sous-tendent.

> COCONSTRUIRE

La coconstruction est une méthode qui consiste à intégrer les citoyennes et les citoyens à l'ensemble du processus d'idéation et de création. Il s'agit d'un processus collaboratif et interactif où les citoyennes et les citoyens ainsi que les intervenantes et les intervenants échangent et créent ensemble.

POURQUOI DÉLIBÉRER SUR L'ÉTHIQUE DE L'IA?

Le déploiement de l'IA a des répercussions sur toutes les sphères de la vie personnelle et sociale; il affecte tout le monde, et personne ne peut mesurer l'ensemble des conséquences de phénomène technologique et social complexe. Il est crucial de multiplier les expertises : les expertises scientifiques, bien sûr, mais aussi les expertises citoyennes, l'expertise des utilisatrices et des utilisateurs de l'IA et des personnes qui en subissent les effets. C'est pourquoi il est essentiel de mettre en œuvre **l'intelligence collective** et d'impliquer le plus grand nombre possible de personnes, au-delà des cercles d'expertes et d'experts et de décideuses et décideurs publics, dans le processus de réflexion sur les questions sociales et éthiques de l'IA.

La délibération permet non seulement d'approfondir nos connaissances de l'IA comme objet technologique qui transforme nos rapports sociaux et politiques, mais elle permet aussi de prendre de meilleures décisions et de conférer à ces décisions une légitimité qui fait souvent défaut aux décisions des expertes et des experts. Cela requiert la participation d'un nombre important de personnes et d'une grande diversité de participantes et de participants. La richesse culturelle et sociale du monde est la seule limite.

Enfin, prendre part aux délibérations permet à chacune et à chacun d'entre vous et d'entre nous de **faire entendre sa voix** et de se donner la chance de mettre l'IA au service du bien commun et de nos intérêts fondamentaux.

NOTIONS ESSENTIELLES

ARGUMENT

Un argument est un raisonnement qui permet de prouver ou de justifier une affirmation (une opinion). Pour défendre une opinion et convaincre avec de bonnes raisons les personnes auxquelles nous nous adressons, nous devons faire appel à un ensemble d'arguments cohérents.

BIEN COMMUN

Le bien commun renvoie à une réalité partagée par toutes et tous, quelle que soit l'organisation sociale. Le « commun » implique l'idée d'un lien entre les membres du groupe. Une innovation pour le bien commun suppose que celle-ci soit développée dans l'intérêt de tout le monde.

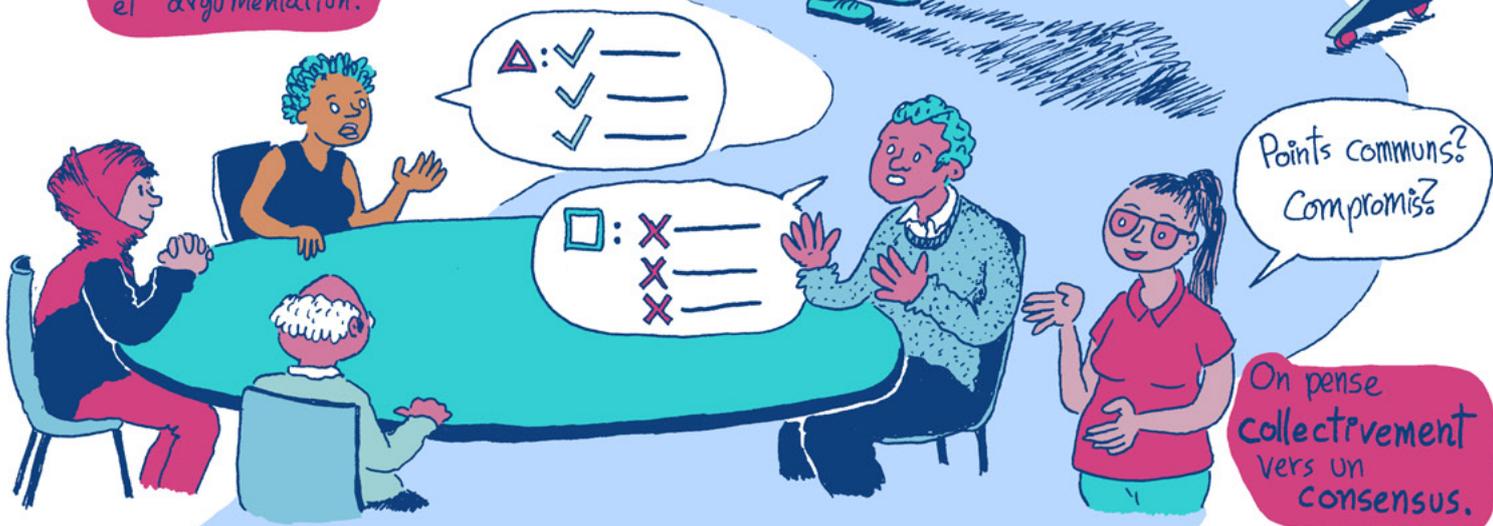
INTELLIGENCE COLLECTIVE

On désigne par l'expression d'intelligence collective la capacité pour un groupe de parvenir à des solutions plus adaptées, de prendre de meilleures décisions et d'approfondir ses connaissances par la discussion, l'échange d'arguments et la mise en commun des connaissances des individus. L'exercice de l'intelligence collective requiert des membres du groupe qu'ils partagent des objectifs et des intérêts communs, ainsi qu'un espace collaboratif (physique ou virtuel).

Une **délibération** regroupe une diversité de personnes ayant différentes opinions sur un sujet donné.



Elle implique une réflexion collective comprenant échange et argumentation.



Un consensus ne veut pas dire que tous et toutes sont d'accord. Un consensus est une position ou une opinion qui apparaît plus raisonnable pour la collectivité.



MartinPM

AFIN DE BIEN DÉLIBÉRER, ON N'OUBLIE PAS QUE...

... les participantes et les participants sont égaux dans la discussion, et la participation à un atelier délibératif requiert un esprit de respect mutuel.

... chaque opinion compte : on ne doit pas exclure sans discussion des opinions soutenues de bonne foi par les participantes et les participants.

... les opinions avancées doivent être soutenues par des arguments et l'échange des arguments doit être public.

CRÉDITS

RÉDACTION

DILHAC, Marc-Antoine

Université de Montréal, OBVIA, Mila-Institut Québécois d'intelligence artificielle, Chaire CIFAR, Algora Lab.

MAI, Vincent

Université de Montréal, Mila-Institut Québécois d'intelligence artificielle, Algora Lab.

MÖRCH, Carl-Maria

Université de Montréal, OBVIA, Algora Lab.

NOISEAU, Pauline

Université de Montréal, OBVIA, Algora Lab.

VOARINO, Nathalie

Université de Montréal, OBVIA, Algora Lab.

ILLUSTRATIONS

PATENAUDE-MONETTE, Martin

martinpm.info

GRAPHISME

HAUSCHILD, Stéphanie

stephaniehauschild.com

CORRECTION DU FRANÇAIS

LACELLE, Pascale

Révidaction

CONTRIBUTIONS DE

FLORES ECHAIZ, Lucia

Université de Montréal, OBVIA, Algora Lab

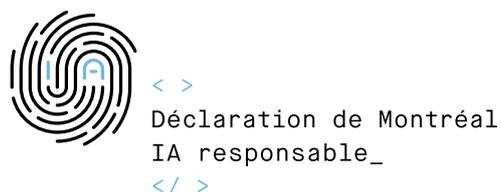
LANTEIGNE, Camylle

Université de Montréal, OBVIA, Algora Lab

SALAZAR GOMEZ, Fatima Gabriela

Université de Montréal, OBVIA, Algora Lab

INSTITUTIONS PARTENAIRES



Merci aux soutiens financiers du gouvernement du Québec, du gouvernement du Canada, du Conseil de recherche en sciences humaines du Canada, des Fonds de recherche du Québec et du Conseil national de recherches du Canada.

