



< >

Déclaration de Montréal IA responsable_

</ >

LA DÉCLARATION DE MONTRÉAL POUR UN DÉVELOPPEMENT RESPONSABLE DE L'INTELLIGENCE ARTIFICIELLE 2018

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](#).

TABLE DES MATIÈRES

LIRE LA DÉCLARATION DE MONTRÉAL	
IA RESPONSABLE	5
PRÉAMBULE	7
<hr/>	
LES PRINCIPES	
1. PRINCIPE DE BIEN-ÊTRE	8
2. PRINCIPE DE RESPECT DE L'AUTONOMIE	9
3. PRINCIPE DE PROTECTION DE L'INTIMITÉ ET DE LA VIE PRIVÉE	10
4. PRINCIPE DE SOLIDARITÉ	11
5. PRINCIPE DE PARTICIPATION DÉMOCRATIQUE	12
6. PRINCIPE D'ÉQUITÉ	13
7. PRINCIPE D'INCLUSION DE LA DIVERSITÉ	14
8. PRINCIPE DE PRUDENCE	15
9. PRINCIPE DE RESPONSABILITÉ	16
10. PRINCIPE DE DÉVELOPPEMENT SOUTENABLE	17
<hr/>	
LEXIQUE	18
<hr/>	
CRÉDITS	21
PARTENAIRES	I

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

LIRE LA DÉCLARATION DE MONTRÉAL IA RESPONSABLE

UNE DÉCLARATION, POUR QUOI FAIRE?

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

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

UNE DÉCLARATION DE QUOI?

DES PRINCIPES

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

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

De ces principes ont été élaborées des recommandations dont l'objectif est de proposer des lignes directrices pour réaliser la transition numérique dans le cadre éthique de la Déclaration. Elles couvrent quelques thèmes intersectoriels clés pour penser la transition vers une société dans laquelle l'IA permet de promouvoir le bien commun : la gouvernance algorithmique, la littératie numérique, l'inclusion numérique de la diversité et la soutenabilité écologique.

UNE DÉCLARATION POUR QUI?

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

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

UNE DÉCLARATION SELON QUELLE MÉTHODE?

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

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

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

APRÈS LA DÉCLARATION?

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

PRÉAMBULE

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

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

Les bénéfices de l'intelligence artificielle seront d'autant plus grands que les risques liés à son déploiement seront faibles. Or le premier danger que présente le développement de l'intelligence

artificielle consiste à donner l'illusion que l'on maîtrise l'avenir par le calcul. Réduire la société à des nombres et la gouverner par des procédures algorithmiques est un vieux rêve qui nourrit encore les ambitions humaines. Mais dans les affaires humaines, demain ressemble rarement à aujourd'hui, et les nombres ne disent pas ce qui a une valeur morale, ni ce qui est socialement désirable.

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

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

1

PRINCIPE DE BIEN-ÊTRE

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

1. Les SIA doivent permettre aux individus d'améliorer leurs conditions de vie, leur santé et leurs conditions de travail.
2. Les SIA doivent permettre aux individus de satisfaire leurs préférences, dans les limites de ce qui ne cause pas de tort à un autre être sensible.
3. Les SIA doivent permettre aux individus d'exercer leurs capacités physiques et intellectuelles.
4. Les SIA ne doivent pas constituer une source de mal-être, sauf si ce dernier permet d'engendrer un bien-être supérieur que l'on ne peut atteindre autrement.
5. L'utilisation des SIA ne devrait pas contribuer à augmenter le stress, l'anxiété et le sentiment de harcèlement liés à l'environnement numérique.



2

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.

PRINCIPE DE RESPECT DE L'AUTONOMIE

1. Les SIA doivent permettre aux individus de réaliser leurs propres objectifs moraux et leur conception de la vie digne d'être vécue.
2. Les SIA ne doivent pas être développés ni utilisés pour prescrire aux individus un mode de vie particulier, soit directement, soit indirectement, en mettant en œuvre des mécanismes de surveillance, d'évaluation ou d'incitation contraignants.
3. Les institutions publiques ne doivent pas utiliser les SIA pour promouvoir ni défavoriser une conception de la vie bonne.
4. Il est indispensable d'encapaciter les citoyens face aux technologies du numérique en assurant l'accès à différents types de savoir pertinents, le développement de compétences structurantes (la littératie numérique et médiatique) et la formation de la pensée critique.
5. Les SIA ne doivent pas être développés pour propager des informations peu fiables, des mensonges et de la propagande, et devraient être conçus dans le but d'en réduire la propagation.
6. Le développement des SIA doit éviter de créer des dépendances par les techniques de captation de l'attention et par l'imitation de l'apparence humaine qui induit une confusion entre les SIA et les humains.

3

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

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

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

4

PRINCIPE DE SOLIDARITÉ

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

1. Les SIA ne doivent pas nuire au maintien de relations humaines affectives et morales épanouissantes, et devraient être développés dans le but de favoriser ces relations et de réduire la vulnérabilité et l'isolement des personnes.
2. Les SIA doivent être développés dans le but de collaborer avec les humains sur des tâches complexes et devraient favoriser le travail collaboratif entre les humains.
3. Les SIA ne devraient pas être mis en œuvre pour remplacer des personnes sur des tâches qui requièrent une relation humaine de qualité, mais devraient être développés pour faciliter cette relation.
4. Les systèmes de santé qui recourent aux SIA doivent prendre en considération l'importance pour les patients des relations avec le personnel médical et la famille.
5. Le développement des SIA ne devrait pas stimuler des comportements cruels avec des robots qui prennent l'apparence d'êtres humains ou d'animaux et semblent agir comme eux.
6. Les SIA devraient permettre d'améliorer la gestion des risques et créer les conditions d'une société de mutualisation des risques individuels et collectifs plus efficace.

5

PRINCIPE DE PARTICIPATION DÉMOCRATIQUE

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



```
class Node {
    constructor(label) {
        this.label = label;
        this.edges = [];
    }
    addEdge(target, weight) {
        const edge = new Edge(this, target, weight);
        this.edges.push(edge);
        target.edges.push(edge);
    }
}

class Edge {
    constructor(source, target, weight) {
        this.source = source;
        this.target = target;
        this.weight = weight;
    }
}

class Graph {
    constructor() {
        this.nodes = [];
    }
    addNode(label) {
        const node = new Node(label);
        this.nodes.push(node);
        return node;
    }
    addEdge(label1, label2, weight) {
        const node1 = this.nodes.find(n => n.label === label1);
        const node2 = this.nodes.find(n => n.label === label2);
        if (!node1 || !node2) {
            throw new Error(`Nodes ${label1} and ${label2} not found`);
        }
        node1.addEdge(node2, weight);
    }
}

const graph = new Graph();
graph.addNode('A');
graph.addNode('B');
graph.addNode('C');
graph.addNode('D');
graph.addNode('E');
graph.addNode('F');
graph.addNode('G');
graph.addNode('H');
graph.addNode('I');
graph.addNode('J');
graph.addNode('K');
graph.addNode('L');
graph.addNode('M');
graph.addNode('N');
graph.addNode('O');
graph.addNode('P');
graph.addNode('Q');
graph.addNode('R');
graph.addNode('S');
graph.addNode('T');
graph.addNode('U');
graph.addNode('V');
graph.addNode('W');
graph.addNode('X');
graph.addNode('Y');
graph.addNode('Z');

graph.addEdge('A', 'B', 1);
graph.addEdge('B', 'C', 1);
graph.addEdge('C', 'D', 1);
graph.addEdge('D', 'E', 1);
graph.addEdge('E', 'F', 1);
graph.addEdge('F', 'G', 1);
graph.addEdge('G', 'H', 1);
graph.addEdge('H', 'I', 1);
graph.addEdge('I', 'J', 1);
graph.addEdge('J', 'K', 1);
graph.addEdge('K', 'L', 1);
graph.addEdge('L', 'M', 1);
graph.addEdge('M', 'N', 1);
graph.addEdge('N', 'O', 1);
graph.addEdge('O', 'P', 1);
graph.addEdge('P', 'Q', 1);
graph.addEdge('Q', 'R', 1);
graph.addEdge('R', 'S', 1);
graph.addEdge('S', 'T', 1);
graph.addEdge('T', 'U', 1);
graph.addEdge('U', 'V', 1);
graph.addEdge('V', 'W', 1);
graph.addEdge('W', 'X', 1);
graph.addEdge('X', 'Y', 1);
graph.addEdge('Y', 'Z', 1);

graph.addEdge('A', 'E', 1);
graph.addEdge('B', 'F', 1);
graph.addEdge('C', 'G', 1);
graph.addEdge('D', 'H', 1);
graph.addEdge('F', 'I', 1);
graph.addEdge('G', 'J', 1);
graph.addEdge('H', 'K', 1);
graph.addEdge('I', 'L', 1);
graph.addEdge('J', 'M', 1);
graph.addEdge('K', 'N', 1);
graph.addEdge('L', 'O', 1);
graph.addEdge('M', 'P', 1);
graph.addEdge('N', 'Q', 1);
graph.addEdge('O', 'R', 1);
graph.addEdge('P', 'S', 1);
graph.addEdge('Q', 'T', 1);
graph.addEdge('R', 'U', 1);
graph.addEdge('S', 'V', 1);
graph.addEdge('T', 'W', 1);
graph.addEdge('U', 'X', 1);
graph.addEdge('V', 'Y', 1);
graph.addEdge('W', 'Z', 1);
```

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

6

PRINCIPE D'ÉQUITÉ

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

1. Les SIA doivent être conçus et entraînés de sorte à ne pas créer, renforcer ou reproduire des discriminations fondées entre autres sur les différences sociales, sexuelles, ethniques, culturelles et religieuses.
2. Le développement des SIA doit contribuer à éliminer les relations de domination entre les personnes et les groupes fondées sur la différence de pouvoir, de richesses ou de connaissance.
3. Le développement des SIA doit bénéficier économiquement et socialement à tous en faisant en sorte qu'il réduise les inégalités et la précarité sociales.
4. Le développement industriel des SIA doit être compatible avec des conditions de travail décentes, et cela, à toutes les étapes de leur cycle de vie, de l'extraction des ressources naturelles jusqu'à leur recyclage, en passant par le traitement des données.
5. L'activité numérique des utilisateurs de SIA et de services numériques devrait être reconnue comme un travail qui contribue au fonctionnement des algorithmes et créé de la valeur.
6. L'accès aux ressources, aux savoirs et aux outils numériques fondamentaux doit être garanti pour tous.
7. Le développement de communs algorithmiques et de données ouvertes pour les entraîner et les faire fonctionner est un objectif socialement équitable qui devrait être soutenu.

7

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.

PRINCIPE D'INCLUSION DE LA DIVERSITÉ

1. Le développement et l'utilisation de SIA ne devraient pas conduire à une uniformisation de la société par la normalisation des comportements et des opinions.
2. Le développement et le déploiement des SIA doivent prendre en considération les multiples expressions des diversités sociales et culturelles, et cela dès la conception des algorithmes.
3. Les milieux de développement de l'IA, aussi bien dans la recherche que dans l'industrie, doivent être inclusifs et refléter la diversité des individus et des groupes de la société.
4. Les SIA doivent éviter d'enfermer les individus dans un profil d'utilisateur ou une bulle filtrante, de fixer les identités personnelles par le traitement des données de leurs activités passées et de réduire leurs options de développement personnel, en particulier dans les domaines de l'éducation, de la justice et des pratiques commerciales.
5. Les SIA ne doivent pas être utilisés ni développés dans le but de limiter la liberté d'exprimer des idées et de communiquer des opinions, dont la diversité est la condition de la vie démocratique.
6. Pour chaque catégorie de service, l'offre de SIA doit être diversifiée afin que des monopoles de fait ne se constituent pas et ne nuisent aux libertés individuelles.

8

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.

PRINCIPE DE PRUDENCE

- 1.** Il est nécessaire de développer des mécanismes qui tiennent compte du potentiel de double-usage (bénéfique et néfaste) de la recherche en IA (qu'elle soit publique ou privée) et du développement des SIA afin d'en limiter les usages néfastes.
- 2.** Lorsque l'utilisation détournée d'un SIA peut représenter un danger sérieux pour la sécurité ou la santé publique, avec une probabilité élevée, il est prudent de restreindre la diffusion publique ou l'accès libre à son algorithme.
- 3.** Avant d'être mis sur le marché, qu'ils soient payants ou gratuits, les SIA doivent satisfaire des critères rigoureux de fiabilité, de sécurité et d'intégrité, et faire l'objet de tests qui ne mettent pas en danger la vie des personnes, ne nuisent pas à leur qualité de vie ni ne portent atteinte à leur réputation ou leur intégrité psychologique. Ces tests doivent être ouverts aux autorités publiques compétentes et aux parties prenantes concernées.
- 4.** Le développement des SIA doit prévenir les risques d'une utilisation néfaste des données d'utilisateurs et protéger l'intégrité et la confidentialité des données personnelles.
- 5.** Les erreurs et les failles découvertes dans les SIA et SAAD devraient être partagées publiquement par les institutions publiques et les entreprises dans les secteurs qui présentent un danger important pour l'intégrité personnelle et l'organisation sociale, et ce, à l'échelle mondiale.

9

PRINCIPE DE RESPONSABILITÉ

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

1. Seuls des êtres humains peuvent être tenus responsables de décisions issues de recommandations faites par des SIA et des actions qui en découlent.
2. Dans tous les domaines où une décision qui affecte la vie, la qualité de la vie ou la réputation d'une personne doit être prise, la décision finale devrait revenir à un être humain et cette décision devrait être libre et éclairée.
3. La décision de tuer doit toujours être prise par des êtres humains et la responsabilité de cette décision ne peut être transférée à un SIA.
4. Les personnes qui autorisent des SIA à commettre un crime ou un délit, ou qui font preuve de négligence en les laissant en commettre, sont responsables de ce crime ou de ce délit.
5. Dans le cas où un tort a été infligé par un SIA, et que le SIA s'avère fiable et a fait l'objet d'un usage normal, il n'est pas raisonnable d'en imputer la faute aux personnes impliquées dans son développement ou son utilisation.



10

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

PRINCIPE DE DÉVELOPPEMENT SOUTENABLE

1. Les équipements de SIA, leurs infrastructures numériques et les objets connectés sur lesquels ils s'appuient comme les centres de données, doivent viser la plus grande efficacité énergétique et minimiser les émissions de gaz à effet de serre (GES) sur l'ensemble de leur cycle de vie.
2. Les équipements de SIA, leurs infrastructures numériques et les objets connectés sur lesquels ils s'appuient, doivent viser à générer un minimum de déchets électriques et électroniques et prévoir des filières de maintenance, de réparation et de recyclage dans une logique d'économie circulaire.
3. Les équipements de SIA, leurs infrastructures numériques et les objets connectés sur lesquels ils s'appuient, doivent minimiser les impacts sur les écosystèmes et la biodiversité à toutes les étapes de leur cycle de vie, notamment lors de l'extraction des ressources naturelles et des étapes de fin de vie.
4. Les acteurs publics et privés doivent soutenir le développement de SIA écologiquement responsables afin de lutter contre le gaspillage des ressources naturelles et des biens produits, de mettre en place des chaînes d'approvisionnement et des échanges commerciaux soutenables, et de réduire la pollution à l'échelle planétaire.

LEXIQUE

Activité numérique

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

Agent conversationnel (*chatbot*)

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

Algorithme

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

Apprentissage machine (*machine learning*)

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

- > En apprentissage supervisé, le système d'intelligence artificielle (SIA) apprend à prédire une valeur à partir d'une donnée entrée. Cela nécessite d'avoir des couples entrée-valeur annotés lors de l'entraînement. Par exemple, un système peut apprendre à reconnaître l'objet présent sur une photo.
- > En apprentissage non-supervisé, le SIA apprend à trouver des similitudes entre des données qui n'ont pas été annotées, par exemple afin de les diviser en différentes partitions homogènes. Ainsi, un système peut reconnaître des communautés d'utilisateurs de réseaux sociaux.

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

Apprentissage profond (*deep learning*)

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

Biens communs numériques (*digital commons*)

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

Bulle de filtre (*filter bubble*)

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

Déconnexion numérique

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

Dépendance de sentier

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

Développement soutenable (sustainable)

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

Données ouvertes (*open data*)

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

Données personnelles

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

Effet rebond

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

Entrainement

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

Fiabilité

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

GAN

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

Intelligence artificielle (IA)

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

Intelligibilité

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

Justifiabilité d'une décision

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

Littératie numérique

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

Soutenabilité écologique forte

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

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

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

Système d'intelligence artificielle (SIA)

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

CRÉDITS

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

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

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

Guillaume Chicoisne, directeur des programmes scientifiques, IVADO

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

NOS PARTENAIRES



CENTRE DE RECHERCHE EN ETHIQUE



ICRA
Programme
IA et
société



Centre interuniversitaire de recherche en analyse des organisations



Canada



Canadian
Cultural
Centre
Paris





< >

Déclaration de Montréal IA responsable_

</ >

declarationmontreal-iaresponsable.com

declarationmontreal-iaresponsable.com