< >

Déclaration de Montréal
IA responsable_

</ >

PARTIE 2

PORTRAIT 2018 DES RECOMMANDATIONS INTERNATIONALES EN ÉTHIQUE DE L'IA

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

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

CHRISTOPHE MONDIN, professionnel de recherche chez CIRANO

GUILLAUME CHICOISNE, directeur des programmes scientifiques, IVADO

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

TABLE DES MATIÈRES

1. INTRODUCTION

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

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

1.1 MÉTHODE

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

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

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

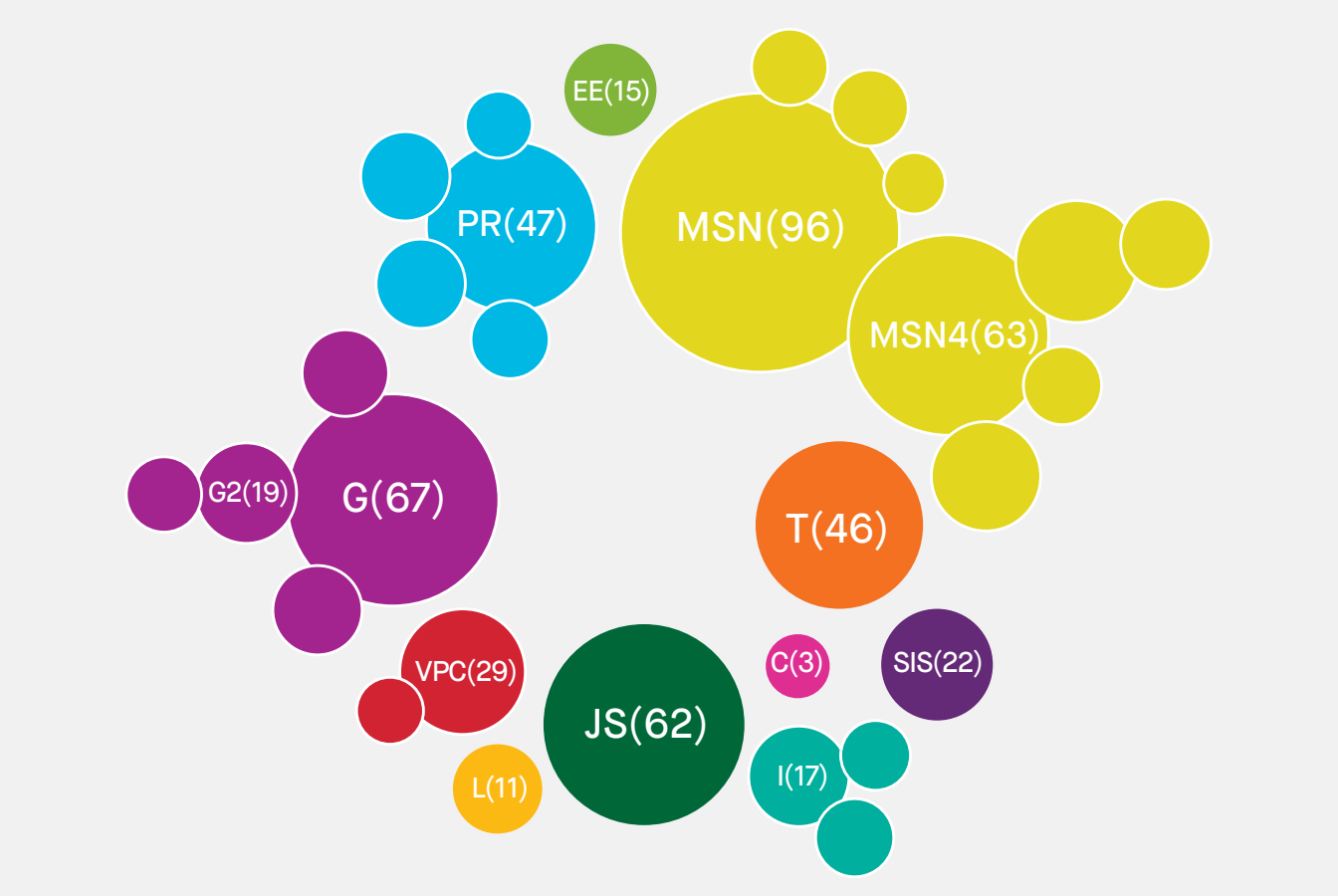
Google, AI Now), tandis que les autres sont beaucoup plus longs et développés, notamment parce qu'ils incluent des considérations d'ordre économiques.

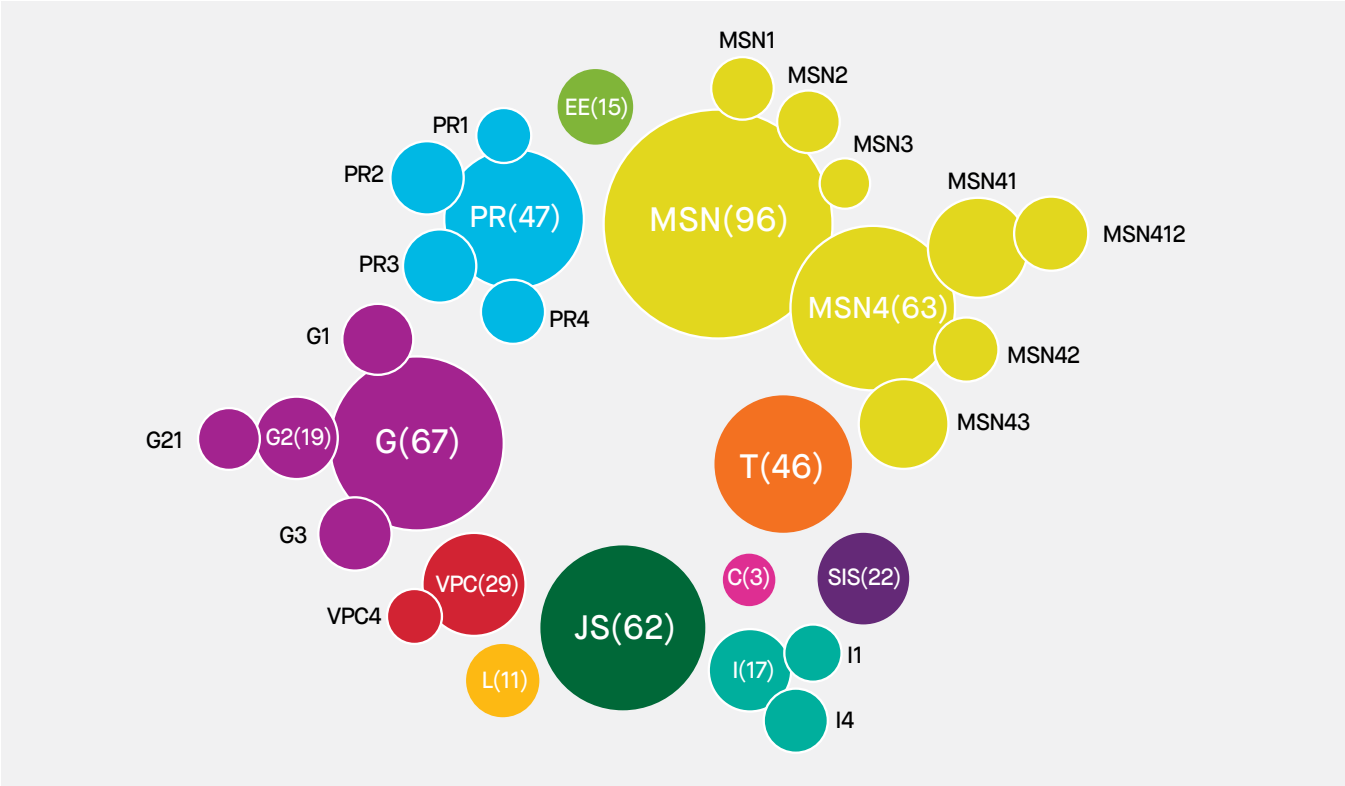
Dans les fiches techniques de la section 3, nous mettons aussi de l'avant la présence, ou non, de principes et de recommandations clairement identifiables. Nous appelons « principes » les propositions très générales, du type « l'IA devrait être bénéfique pour la société » tandis que les « recommandations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

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

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

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





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

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

1.2

REMARQUES LIMINAIRES

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

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

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

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

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

2. SYNTHÈSE THÉMATIQUE DES RECOMMANDATIONS

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

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

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

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

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

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

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

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

BIEN-ÊTRE

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

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

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

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

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

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

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

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

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

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

AUTONOMIE

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

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

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

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

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

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

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

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

JUSTICE

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

VIE PRIVÉE

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

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

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

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

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

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

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

CONNAISSANCE

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

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

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

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

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

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

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

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

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

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

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

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

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

DÉMOCRATIE

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

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

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

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

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

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

RESPONSABILITÉ

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.1 LES SEPT RAPPORTS RETENUS

(AI NOW) AI NOW 2017 REPORT

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

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

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

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

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

(ASILOMAR) ASILOMAR AI PRINCIPLES

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

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

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

(VILLANI) DONNER UN SENS À L'INTELLIGENCE ARTIFICIELLE

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

(GOOGLE) AI AT GOOGLE : OUR PRINCIPLES

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

3.2
RAPPORTS EXAMINÉS, MAIS NON RETENUS

A NEXT GENERATION ARTIFICIAL INTELLIGENCE DEVELOPMENT PLAN

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

STRATEGY FOR DENMARK'S DIGITAL GROWTH

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

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

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

FINLAND'S AGE OF ARTIFICIAL INTELLIGENCE

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

ETHICS COMMISSION AUTOMATED AND CONNECTED DRIVING

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

NATIONAL STRATEGY FOR ARTIFICIAL INTELLIGENCE #AIFORALL

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

ARTIFICIAL INTELLIGENCE AT THE SERVICE OF CITIZENS

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

ARTIFICIAL INTELLIGENCE TECHNOLOGY STRATEGY

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

TOWARDS AN AI STRATEGY IN MEXICO

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

SHAPING A FUTURE NEW ZEALAND

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

ARTIFICIAL INTELLIGENCE IN SWEDISH BUSINESS AND SOCIETY

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

INDUSTRIAL STRATEGY

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

PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE

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

THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN

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

ARTIFICIAL INTELLIGENCE, AUTOMATION, AND THE ECONOMY

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

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

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

3.3 AUTRES RAPPORTS CONSULTÉS

(SUÈDE) NATIONAL APPROACH FOR ARTIFICIAL INTELLIGENCE

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

(ALLEMAGNE) ECKPUNKTE DER BUNDESREGIERUNG FÜR EINE STRATEGIE KÜNSTLICHE INTELLIGENZ

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

(FINLANDE) WORK IN THE AGE OF ARTIFICIAL INTELLIGENCE

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

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

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

(AUSTRALIE) AUSTRALIA 2030: PROSPERITY THROUGH INNOVATION

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

Merci à Paloma Fernandez-McAuley pour son aide.

CRÉDITS DU RAPPORT FINAL

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

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

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

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

Coordination

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

Collaboration aux contenus

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

Révision et édition

Chantal Berthiaume, gestionnaire de contenu et rédactrice

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

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

Traduction

Rachel Anne Normand et François Girard, Services linguistiques Révidaction

Graphisme

Stéphanie Hauschild, directrice artistique

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

NOS PARTENAIRES

Université de Montréal

CRE
CENTRE DE RECHERCHE EN ETHIQUE

LAB VILLE PROSPECTIVE

IVADO
INSTITUT DE VALORISATION DES DONNÉES

SAT

Mila

ICRA
Programme
IA et
société

Université de Montréal | designNsociété

polEtHics
CHAIRE DE RECHERCHE DU CANADA
ÉTHIQUE PUBLIQUE ET THÉORIE POLITIQUE

CIRANO
Allier savoir et décision
Centre interuniversitaire de recherche en analyse des organisations

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

UNIVERSITÉ LAVAL
Institut d'éthique appliquée

MUSÉE DE LA CIVILISATION
Québec

CENTRE D'ÉTUDES ET DE RECHERCHES INTERNATIONALES
CÉRIUM
Université de Montréal

crdm_ul
CENTRE DE RECHERCHE EN DONNÉES MASSIVES DE L'UNIVERSITÉ LAVAL

Canada

Centre Culturel Canadien Paris
Canadian Cultural Centre Paris

MEC
Maison des étudiants canadiens

< >

Déclaration de Montréal
IA responsable_

</ >

declarationmontreal-iaresponsable.com