

Vers un développement responsable de l'IA

Soirée de réflexion autour de la Déclaration de Montréal



ESG UQÀM

UQÀM | **Sciences**

UQÀM | **Sciences humaines**

UQÀM | **Science politique et droit**

Compte-rendu en vue de la participation à la campagne de consultation de la
Déclaration de Montréal pour un développement responsable de l'intelligence artificielle

Mars 2018

UQÀM
Université du Québec à Montréal

Contact et organisation :

Dominic Martin
Professeur au Département d'organisation et ressources humaines
Université du Québec à Montréal
C.P. 8888, succ. Centre-ville
Montréal QC H3C 3P8
Canada

Sébastien Gambs
Professeur au Département d'informatique
Université du Québec à Montréal

Nous remercions les personnes suivantes pour leur collaboration à la soirée de réflexion et à la production de ce compte-rendu.

Magali Barré
Agente de recherche et de planification
Université du Québec à Montréal

Yoshua Bengio
Professeur au Département d'informatique et de
recherche opérationnelle
Université de Montréal

Chloé T. Bergeron
Étudiante à la maîtrise en gestion
Université du Québec à Montréal

Josée Boileau
Journaliste et auteure

Maude Bonenfant
Professeur au Département de communication
sociale et publique
Université du Québec à Montréal

René Canuel
Agent de recherche et de planification
Université du Québec à Montréal

Hugo Cyr
Doyen de la Faculté de sciences politiques et de droit
Université du Québec à Montréal

Martin Gibert
Agent de recherche
Centre de recherche en éthique / IVADO
Université de Montréal

Étienne Harnad
Professeur au Département de psychologie
Université du Québec à Montréal

Antoine Laurent
Étudiant à la maîtrise en informatique
Université du Québec à Montréal

Christophe Malaterre
Professeur au Département de philosophie
Université du Québec à Montréal

Marie-Jean Meurs
Professeure au Département d'informatique
Université du Québec à Montréal

Nicolas Riendeau
Agent de recherche et de planification
Université du Québec à Montréal

Josée Savard
Agente de recherche et de planification
Université du Québec à Montréal

Geneviève Shields
Agente de recherche et de planification
Université du Québec à Montréal

Avant-propos	3
I. Résumé et principales recommandations.....	3
II. Présentation de la Déclaration	4
III. Commentaires sur la Déclaration	5
IV. Aperçu des interventions du public	8

Avant-propos

Le jeudi 15 février 2018 à l'Agora Hydro-Québec de l'Université du Québec à Montréal (UQAM) s'est tenue une soirée de réflexion autour de la *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*ⁱ. La première version de la Déclaration a été présentée en novembre 2017. Il s'en est suivi une période de consultation sur cette première version qui se termine à la fin du mois de mars 2018. Dans la mouvance de cette consultation, l'UQAM a organisé un événement s'intitulant «Pour un développement responsable de l'IA : Soirée de réflexion autour de la Déclaration de Montréal»ⁱⁱ. En effet, on anticipe des transformations sociales importantes en lien avec l'IA et on reconnaît l'importance de réfléchir à ces enjeux. Afin d'apporter un éclairage multidisciplinaire sur le sujet, la soirée visait à réunir des chercheurs de l'UQAM provenant de différentes disciplines autour des questions soulevées par la Déclaration et à participer à cette campagne de coconstruction. En particulier, les participants ont été invités à préciser pourquoi les principes de la Déclaration sont importants, quelles en sont les grandes implications éthiques et sociétales, comment devrait-on mettre la Déclaration en pratique, ainsi que toute autre considération jugée pertinente.

L'événement a été organisé par L'École des sciences de la gestion, la Faculté des sciences, la Faculté des sciences humaines et la Faculté de science politique et droit de l'UQAM, en collaboration avec l'Institut d'études internationales de Montréal. Il visait à rejoindre la communauté universitaire dans son ensemble. Plus de 200 professeurs, étudiants, partenaires du milieu des affaires ou autres membres de la communauté étaient présents.

Durant la soirée, chaque chercheur de l'UQAM a été associé à un principe de la Déclaration. Yoshua Bengio et Martin Gibert sont venus présenter la déclaration dans son ensemble et Josée Boileau était chargée de l'animation. Deux périodes de questions ont permis aux membres du public de faire part de leurs interrogations ou réflexions. Pour ce faire, ils devaient utiliser l'adresse courriel soireeia@uqam.ca ou envoyer directement leur intervention sur la plateforme Twitter par l'entremise du mot-clic [#soireeia](https://twitter.com/soireeia). Le Tableau 1 donne un aperçu du déroulement de l'événement.

Le présent rapport offre un aperçu des commentaires des participants et des interventions du public. Il sera déposé au comité directeur de la Déclaration dans le cadre de la campagne de consultation.

I. Résumé et principales recommandations

Un large consensus s'est dégagé au cours de la soirée sur l'importance de réfléchir sur les impacts sociaux de l'IA. Ces nouveaux développements laissent entrevoir des transformations majeures de la société et on doit s'attaquer à ces enjeux dès maintenant. On évoque notamment la possibilité d'une croissance économique et de grands progrès scientifiques et sociaux, l'automatisation de certaines formes de travail peu valorisantes, et la possibilité d'un apprentissage personnalisé. Cependant, on met aussi en garde contre les dangers potentiels de ces nouvelles technologies : créations d'outils de surveillance et de contrôle, chômage technologique, accroissement des inégalités économiques et politiques, discrimination ou création de nouvelles armes.

Tableau 1 — Déroulement de la soirée

Animation

Josée Boileau

Présentation de la Déclaration (approx. 20 min.)

Yoshua Bengio

Martin Gibert

Commentaires sur la Déclaration (approx. 110 min.)

Marie-Jean Meurs : Bien-être

Christophe Malaterre : Autonomie

Hugo Cyr : Justice

Période de questions (20 min.)

Sébastien Gambs : Vie privée

Étienne Harnad : Connaissance

Dominic Martin : Démocratie

Maude Bonenfant : Responsabilité

Période de questions (20 min.)

La Déclaration a été créée dans le but de susciter un débat public et de proposer une orientation progressiste et inclusive du développement de l'IA. Parmi les principaux commentaires sur la Déclaration se dégageant de la soirée, on retrouve les recommandations suivantes :

- Le développement de l'IA peut être harmonieux s'il est encadré par des politiques publiques éclairées.

- Il faut s'assurer que le développement de l'IA ne remplace pas la prise de décision par des êtres humains responsables et imputables, en particulier lorsqu'il s'agit de décisions pouvant avoir un impact important sur les individus.

ⁱ <https://www.declarationmontreal-iaresponsable.com/>

ⁱⁱ <https://recherche.esg.uqam.ca/reflexion-declaration-mtl-ia/>

– Le développement de l’IA doit permettre aux individus de contrôler la collecte et l’usage de leurs données personnelles.

– On doit faire mention directement de la notion de justice distributive et du respect de l’égalité humaine dans la déclaration.

– Il faut garantir la transparence sur les données et les algorithmes afin de pouvoir assurer la responsabilité algorithmique et la non-discrimination.

– En plus de l’importance accordée au bien-être et au jugement des êtres humains, il faut inclure tous les êtres sentients.

– Il faut rendre explicites les types d’autonomie (autonomie morale, fonctionnelle et politique; et autonomie des êtres humains ou des IA) que l’on veut promouvoir face aux risques et incertitudes associés aux systèmes intelligents.

– On doit éviter la concentration du pouvoir politique et économique liée au développement et à l’exploitation de l’IA.

– Il ne faut pas que quelques entreprises soient les seuls à tirer des bénéfices importants de l’IA tout en laissant aux gouvernements la responsabilité de pallier les coûts qui seront associés aux erreurs et aux abus.

– On doit compenser les nouvelles inégalités économiques et politiques en distribuant le capital de manière plus égalitaire et en encourageant la création de nouvelles structures de propriété d’entreprises telle que les coopératives.

– La communauté scientifique a la responsabilité de soutenir le débat, d’expliquer et de dégager les nouveaux enjeux liés au développement technologique. Les citoyens ont aussi la responsabilité de s’informer et de se questionner sur l’introduction de l’IA dans leur vie quotidienne.

II. Présentation de la Déclaration

Présentation Yoshua Bengio

La présentation de **Yoshua Bengio** avait pour but de lancer la réflexion sur le développement de l’IA. Il a commencé par rappeler l’importance d’avoir une discussion collective sur ces questions, car il entrevoit des impacts et enjeux sociétaux importants. Il s’intéresse aussi aux espoirs qu’apporte l’IA pour ensuite s’attaquer aux dangers sous-jacents.

Concernant les espoirs, M. Bengio anticipe une croissance économique importante à la suite du développement de l’IA dans les prochaines années (en lien notamment avec la vallée de l’IA à Montréal). Cette croissance peut mener à des progrès sociaux et technologiques importants ainsi que des avancées en lien avec le domaine de la santé, de l’éducation et de la justice. Il est aussi possible que l’IA nous libère complètement de certaines formes de travail, ce qui pourrait amener à une

réduction des emplois répétitifs ainsi qu’un plus grand épanouissement dans le travail en général.

M. Bengio évoque l’amélioration de la possibilité d’espionner des personnes comme premier danger potentiel de l’IA, par exemple grâce aux algorithmes de reconnaissance faciale. La reconnaissance faciale pourrait aussi être mise au service d’applications militaires, tels des robots tueurs, c’est-à-dire des robots capables de prendre par eux-mêmes la décision de tuer un être humain. Aussi, il souligne qu’on pourrait revenir sur la proposition faite selon laquelle le développement apporterait une meilleure relation entre être humain et travail. L’automatisation risque d’avoir un effet, au moins temporaire, sur le marché de l’emploi. Un autre danger serait l’utilisation d’IA pour produire de la publicité ciblée. En effet, la publicité, qui peut être vue comme une manipulation de l’esprit, peut avoir un impact important si elle est personnalisée (par exemple dans le contexte politique).

M. Bengio termine sa présentation avec l’idée qu’on pourrait voir un renforcement de la discrimination sociale causée par les biais présents dans les données d’apprentissages qui sont nécessaires à l’IA, ainsi que par la concentration du savoir et de la richesse.

Présentation Martin Gibert

Martin Gibert est le coordinateur de la Déclaration. L’objectif principal de son intervention était de bien faire comprendre au public les buts et le contenu du travail fait jusqu’à présent, ainsi que les impacts de l’IA sur la société. La Déclaration a été lancée le 3 novembre à Montréal au palais des congrès à l’occasion d’un forum sur l’intelligence artificielle responsable organisé par l’Université de Montréal. Le but de cette initiative est de susciter un débat public et de proposer une orientation progressiste et inclusive du développement de l’IA. La Déclaration est développée selon un processus de coconstruction afin d’être enrichie par des propositions concrètes. En effet, chacun peut participer à la construction de la Déclaration de plusieurs manières : questionnaire en ligne, dépôt d’un mémoire ou participation à des événements comme des journées de coconstruction et des cafés citoyens.

M. Gibert rappelle sommairement le contenu de la déclaration. Elle est composée d’un préambule qui contient les définitions des principaux concepts de l’IA responsable, telle que par exemple la définition suivante : **« Qu’elle soit naturelle ou artificielle, l’intelligence n’a pas de valeur en soi. L’intelligence d’un individu ne nous dit rien de sa valeur morale; c’est aussi le cas pour toute autre entité intelligente. L’intelligence peut néanmoins avoir une valeur instrumentale : c’est un outil qui peut nous éloigner ou nous rapprocher d’un objectif que nous valorisons »**. Les valeurs fondamentales composant cette déclaration sont : le **bien-être**, l’**autonomie**, la **justice**, la **vie privée**, la **connaissance**, la **démocratie** et la **responsabilité**. Chacune

de ces valeurs est composée d'un principe et centrée autour de cinq questions qui favorisent la réflexion. Il termine en suggérant que l'IA nous met face à un nouveau domaine et on peut avoir l'espoir de faire des IA qui respectent les principes éthiques.

III. Commentaires sur la Déclaration

BIEN-ÊTRE
Marie-Jean
Meurs

Le commentaire de **Marie-Jean Meurs** portait sur le principe du **bien-être**, dont la formulation proposée est la suivante : « **Le développement de l'IA devrait ultimement viser le bien-être de tous les êtres**

sentients ». Elle traite cette proposition en fonction des connaissances qu'elle possède sur l'intelligence artificielle, la modélisation mathématique et l'application des algorithmes d'apprentissage sur des données réelles. Elle choisit ensuite de discuter de six domaines dans lesquels l'IA aura un impact tangible : la santé, la sécurité, l'environnement, le travail, l'éducation et la justice.

Concernant le domaine de la santé, elle souligne que l'utilisation de l'IA n'est pas récente, mais elle s'améliore grâce à la disponibilité de données massives, la pertinence d'algorithmes de plus en plus développés et la puissance de calcul plus importante. Il est désormais possible de faire de la médecine personnalisée, grâce à l'analyse de grands volumes de données médicales, qui permet de meilleures décisions, des traitements autoadaptatifs et de la détection précoce des risques.

En matière de sécurité, les avancées en IA ont permis d'améliorer la détection d'activités suspectes et de découvrir des données sensibles parmi une grande masse de données. Ces technologies, qui contribuent à améliorer la sécurité des citoyens, peuvent également être utilisées pour protéger leur vie privée en détectant les fraudes et les fuites de données. L'IA a aussi son lot d'applications en matière de protection de l'environnement. Une de ces applications est la surveillance automatique des territoires qui permet entre autres d'intervenir contre l'orpaillage sauvage, en Amazonie par exemple, ou de détecter les activités de pêche illégale. Les nombreuses applications de l'IA à la génomique contribuent à faire avancer les connaissances tant en santé humaine et animale qu'en optimisation des ressources naturelles et agricoles.

En terminant, Mme Meurs illustre le fait que l'on commence à rechercher des méthodes pour faire de l'apprentissage personnalisé en éducation, ce qui rejoint l'idée de Yoshua Bengio concernant l'automatisation du travail. Enfin, elle souligne le fait que les biais ne sont pas présents dans les algorithmes, mais dans les données. Elle répond de façon optimiste à la question : « Peut-on mettre en relation IA et bien-être ? » Oui le développement de l'IA peut être harmonieux s'il est encadré par des politiques publiques éclairées.

AUTONOMIE
Christophe
Malaterre

Le commentaire de **Christophe Malaterre** portait sur le principe **autonomie**, dont la formulation proposée est la suivante : « **Le développement de l'IA devrait favoriser l'autonomie de tous les êtres humains et contrôler, de manière responsable, celle des systèmes informatiques** ». Associer autonomie et intelligence artificielle (IA) peut se comprendre de différentes façons selon lui en raison de la définition de l'autonomie qu'on adopte et le sujet à qui cette autonomie est attribuée. Étymologiquement, autonomie provient du grec autos (soi-même) et nomos (loi, règle). L'autonomie se comprend alors comme « la condition d'une entité qui choisit elle-même les lois auxquelles elle se soumet ». Cette définition soulève aussitôt deux questions : « l'autonomie de qui ? » et « une autonomie vis-à-vis de quelles lois ? ».

L'autonomie est souvent comprise comme portant sur des lois qui peuvent être ou bien politiques, ou bien morales. Dans le premier cas, souvent considéré au niveau collectif, l'autonomie correspond à la capacité d'un groupe à s'organiser et à s'administrer lui-même. Dans le second, plutôt individuel, l'autonomie caractérise la condition d'un individu qui détermine lui-même les lois morales auxquelles il se soumet. L'étymologie du mot autonomie nous renvoie donc à deux types d'autonomie : (a) une « autonomie politique » et (b) une « autonomie morale ».

Mais il est intéressant de noter que la notion d'autonomie peut aussi être comprise en un sens plus éloigné de ses racines étymologiques, comme c'est tout particulièrement le cas lorsqu'on oppose autonomie à dépendance. L'autonomie devient alors comprise comme la capacité de fonctionner ou de « se débrouiller » seul. C'est en ce sens qu'on peut parler de l'autonomie sociale d'un individu comme étant la capacité de cet individu à s'intégrer par lui-même dans un groupe social, ou encore de l'autonomie médicale d'un individu, notamment dans le cas d'une personne âgée ou en situation de handicap. M. Malaterre propose de qualifier ce troisième type d'autonomie comme étant (c) une « autonomie fonctionnelle ».

Indépendamment du type d'autonomie en jeu, il est important de préciser aussi le sujet de cette autonomie. Dans le contexte de l'IA, l'autonomie peut être (i) celle des êtres humains lorsqu'ils sont aidés par des systèmes IA. Mais elle peut aussi concerner (ii) celle des êtres humains (et, le cas échéant, des autres organismes biologiques) dans un environnement désormais peuplé de systèmes IA. A contrario, cette autonomie peut être comprise comme étant (iii) l'autonomie des systèmes IA dans un environnement humain.

Croiser les trois formes d'autonomie (a-c) et les trois sujets possibles de cette autonomie (i-iii) produit une grille d'analyse de neuf manières différentes de comprendre la notion d'autonomie dans le contexte de l'IA. On peut ainsi évoquer, par exemple, l'autonomie fonctionnelle des êtres

humains aidés par des systèmes IA (c+i), l'autonomie politique des humains dans un monde peuplé de systèmes IA (a+ii) ou encore l'autonomie morale des systèmes IA (b+iii).

Selon M. Malaterre, la formulation proposée dans la Déclaration fait l'impasse sur plusieurs de ces manières de concevoir l'autonomie. Elle met l'accent sur une autonomie des êtres humains aidés par des systèmes IA de type (c+i), tout en évoquant une nécessité de contrôle des systèmes IA, renvoyant ainsi peut-être à des formes d'autonomie de type (b+iii) ou (c+iii). Il propose de rendre plus explicites, dans la Déclaration, les types d'autonomie ciblés par le principe d'autonomie et, face aux risques et incertitudes associés aux systèmes IA, de renforcer ce principe pour le rendre plus volontariste. Il propose ainsi une reformulation du principe d'autonomie selon laquelle «Les systèmes IA ne doivent pas nuire à l'autonomie (morale, fonctionnelle, politique) des êtres humains, mais devraient chercher à y contribuer. Les systèmes IA ne doivent pas être rendus entièrement autonomes des êtres humains, mais doivent demeurer sous leur contrôle (moral, fonctionnel, politique)».

JUSTICE Hugo Cyr

Le commentaire d'Hugo Cyr portait sur le principe **justice**, dont la formulation proposée est la suivante : «**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**». M. Cyr nous propose de nous questionner sur le type de droit dont chacun peut se prévaloir.

Selon Aristote, la justice se classifierait en différentes catégories. La justice commutative représente l'égalité arithmétique des prestations. En revanche, la juste valeur de l'échange dépend d'un ensemble de faits, telles l'égalité de l'information, la présence de coercition... La justice peut intervenir pour tenter de rééquilibrer les déséquilibres lors des échanges.

Une autre catégorie est celle de la justice distributive, qui est la répartition des biens en proportion de certains critères de mérite. Les critères de mérite font l'objet de controverses. En effet, utiliser la «dignité humaine» comme critère de mérite favoriserait l'identité de traitement pour assurer l'égale dignité. Avec ce critère, on pourrait justifier la différence de traitement pour rétablir une situation factuelle qui ne correspondrait pas à l'égale dignité. M. Cyr nous propose de placer ce concept au cœur de la conception de la justice qui sera adoptée concernant l'IA. L'exigence de la reconnaissance de l'égale dignité signifie que les décisions de justice devront ultimement être imputables à un humain responsable.

En terminant sa présentation, M. Cyr propose quelques

critiques de la formulation actuelle du principe justice. Premièrement, pour éviter que seule la justice commutative ne soit prise en compte dans la proposition sur la justice, il conseille de faire mention explicitement de la notion de justice distributive. Deuxièmement, l'objectif de respect de l'égale dignité humaine devrait être mentionné. Ce principe pourrait inclure le fait que certaines décisions doivent être prises par des humains responsables et imputables, et non par des machines. Troisièmement, le passage sur la discrimination est une illustration d'un échec de la justice commutative et de la justice distributive et non un principe égal à celui de la promotion de la justice. Une reformulation serait nécessaire.

La proposition de reformulation de M. Cyr se traduirait donc par : «Le développement de l'IA devrait promouvoir la justice sociale et respecter l'égale dignité humaine, notamment en visant à éliminer toute forme de discrimination, incluant celle liée au genre, l'âge... et aux origines ethniques, aux statuts sociaux».

VIE PRIVÉE Sébastien Gambs

Le commentaire de **Sébastien Gambs** portait sur le principe **vie privée**, dont la formulation proposée est la suivante : «**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**». M. Gambs mentionne que les données massives sont le carburant faisant avancer l'intelligence artificielle et que la plus grande partie de celles-ci sont constituées de données personnelles, ce qui peut créer de graves problèmes en termes de respect de la vie privée. Une solution à cette problématique pourrait être l'anonymisation des données, mais cette approche n'est pas suffisamment mature pour être utilisable à l'heure actuelle.

Selon M. Gambs, si on laisse les algorithmes faire des choix sans qu'on puisse comprendre le raisonnement derrière ceux-ci, on finira par ne plus être en mesure d'avoir des explications sur leur prise de décision, et donc indirectement les rendre imputables par rapport à cette décision. Il faudra donc travailler à une amélioration des explications de leurs prises de décision.

Deux éléments principaux contribuent à l'accroissement des risques de perte de vie privée. Premièrement, la capacité à enregistrer et stocker les informations personnelles a augmenté de manière régulière dans les dernières années. Deuxièmement, le mouvement d'ouverture des données conduit à la divulgation d'énormes quantités de données.

Dans le futur, il sera donc nécessaire d'améliorer la transparence de ces données. Cela pourra être fait par des approches de type législation, par des efforts faits par les acteurs académiques ou industriels pour l'instauration de

la transparence dans la conception des algorithmes, par la publication des codes sources appris par l'IA, par l'utilisation de modèles interprétables ou par l'étude des biais.

La transparence est une première étape, mais elle ne doit pas être le but final, selon M. Gambs. On devrait plutôt viser une responsabilité algorithmique, c'est-à-dire d'avoir des outils/méthodes qui puissent vérifier que l'IA qui a été mise en place respecte certaines normes éthiques de la société. Du côté des ingénieurs, il y aurait des recherches à faire sur les méthodes de validations de propriétés éthiques d'une IA.

En conclusion, tous les risques de vie privée qui existaient auparavant deviennent de plus en plus importants. Il faudra donc trouver un équilibre entre les bénéfices sociétaux et économiques de l'IA, tout en garantissant le respect de la vie privée et des autres droits fondamentaux des individus.

La proposition de reformulation de M. Gambs pourrait se traduire de la manière suivante : « Le développement de l'IA doit se faire dans le respect des droits fondamentaux tel que le droit à la protection de la vie privée, en permettant en particulier un contrôle fort des individus sur la collecte et l'usage de leurs données personnelles. De plus, la transparence sur les données et les algorithmes mis en œuvre doit être garantie afin de pouvoir assurer une responsabilité algorithmique de l'IA garantissant que son développement est éthiquement aligné sur des principes fondamentaux telle que la non-discrimination. »

CONNAISSANCE Étienne Harnad

Le commentaire d'Étienne Harnad portait sur le principe **connaissance**, dont la formulation proposée est la suivante : « **Le développement de l'IA devrait promouvoir la pensée critique et nous prémunir contre la propagande et la manipulation** ». Avant l'ère de l'ordinateur, les seuls titulaires de l'intelligence étaient les êtres vivants. M. Harnad nous propose de voir l'intelligence générale, qu'elle soit naturelle ou artificielle, comme étant le mécanisme sous-jacent à notre capacité d'action. Tout ce que nous pouvons faire est généré par un mécanisme qui est l'incorporation de l'intelligence.

Lorsqu'on parle d'IA en sciences cognitives, le but est d'arriver à un point où l'on pourrait créer une IA capable de faire tout ce que peut faire une intelligence naturelle, à tel point qu'il ne serait pas possible de les distinguer (le Test de Turing). Si cette situation se présente, on pourrait dire qu'on a réussi d'une certaine manière à bien comprendre l'intelligence naturelle.

Une machine est un système causal. Si on y réfléchit, un être humain est lui aussi un système causal. Ainsi, les machines sont construites par des êtres humains et reproduisent synthétiquement beaucoup de choses. Mais M. Harnad pose la question : peut-on dire que les IA ne font pas de vraies choses parce qu'elles agissent

artificiellement? Est-ce que seul le jugement humain devrait pouvoir prendre des décisions? Le jugement humain fait des erreurs. Nous prenons tous des décisions basées sur des algorithmes pour faire des choix, tout comme les (autres) machines le font (en se basant sur ses connaissances, ses expériences). Il est donc essentiel de se poser la question : au jugement de qui veut-on se fier? Ainsi, est-ce qu'on devrait par exemple se fier seulement à un humain parce qu'il saigne? Cependant, si on vise le bien-être des êtres qui saignent, on devrait viser celui de tous les êtres vivants, incluant celui des animaux.

DÉMOCRATIE Dominic Martin

Le commentaire de **Dominic Martin** portait sur le principe **démocratie**, dont la formulation proposée est la suivante : « **Le développement de l'IA devrait favoriser la participation éclairée à la vie publique, la coopération et le débat démocratique** ». Devant cette proposition, il se demande d'abord quels sont les risques que pose l'IA pour la démocratie et il identifie cinq risques importants. Premièrement, l'IA peut avoir des effets épistémiques potentiellement néfastes. Deuxièmement, on peut faire face à des problèmes d'éthique des algorithmes. Troisièmement, il y a des problèmes liés à la compétence morale des systèmes IA et au niveau de confiance qu'on doit accorder à la prise de décision morale de ces systèmes. Quatrièmement, il se préoccupe du pouvoir économique et politique croissant des grandes entreprises de technologie et, cinquièmement, des problèmes liés à la croissance des inégalités économiques.

Son commentaire traite plus spécifiquement des deux derniers problèmes, soit les problèmes liés aux grandes entreprises de technologie et ceux liés aux inégalités économiques. Ainsi, le pouvoir grandissant des grandes entreprises de technologie est préoccupant pour plusieurs raisons. Ces entreprises ont des ressources économiques extrêmement considérables. De plus, elles ont des liens étroits avec le pouvoir politique et ont accès à un volume de données personnelles très important. Elles sont chefs de file en matière du développement de l'IA et de ses applications commerciales. Et, finalement, elles ont souvent des ambitions commerciales monumentales.

Concernant le problème des inégalités économiques, il souligne que depuis les années 1950, les inégalités de revenu n'ont fait que croître. Le développement de l'IA exacerbe ce problème, en plus de créer un nouvel effet pervers. En effet, lorsque l'on automatise la production, on se retrouve à déplacer la richesse du revenu vers le capital puisqu'on donne moins d'argent à des êtres humains pour leur travail et on utilise cette richesse pour développer de l'infrastructure, des logiciels, etc. Or, s'il y a présence de problèmes en termes d'inégalité de revenus dans notre société, les problèmes d'inégalité en termes de capital sont encore plus grands, puisque le capital y est beaucoup plus concentré que le revenu.

Il termine son commentaire en proposant un certain nombre de solutions comme briser les grands monopoles dans l'industrie technologique. On peut aussi tenter d'encadrer les activités de ces entreprises. Un revenu universel ou un revenu minimum garanti est une autre option intéressante. Ce revenu peut être financé à partir d'une taxe sur l'automatisation ou d'une taxe sur le capital, la deuxième forme de taxe étant la plus prometteuse. Finalement on peut s'attaquer plus directement à la source du problème en distribuant le capital de manière plus égalitaire et en encourageant de nouvelles structures de propriété d'entreprises telle que les coopératives.



Le commentaire de **Maude Bonenfant** portait sur le principe **responsabilité**, dont la formulation proposée est la suivante : « **Les différents acteurs du développement de l'IA devraient assumer leur responsabilité en œuvrant contre les risques de ces innovations technologiques** ». La responsabilité peut se définir comme une obligation légale ou une nécessité morale et intellectuelle de réparer une faute, de remplir un devoir, une charge ou un engagement. Pour donner suite à cette définition, une question s'impose : qui sera responsable de l'IA ? Par exemple, son développement pourrait aller jusqu'à l'utilisation de robots tueurs, des armes dont l'usage sera entièrement automatisé. D'ailleurs, des chercheurs en IA ont signé récemment une lettre ouverte demandant l'interdiction des armes autonomes puisque, dans ce contexte, on se demande qui deviendrait responsable.

Mme. Bonenfant présente une double dynamique des limites et du pouvoir liés à la responsabilité. Quel type de pouvoir veut-on céder à l'IA et que veut-on conserver ? En lui cédant du pouvoir, est-ce que l'on se déresponsabilise ? Pour répondre à ce problème, il est nécessaire que l'ensemble des acteurs sociaux se mobilise et prenne des décisions visant d'abord le bien-être collectif.

En effet, des bénéfices monétaires seront tirés de l'exploitation de l'IA, principalement par les entreprises. En revanche, il faudra aussi que des coûts y soient associés. Il ne faudrait pas que quelques entreprises tirent de grands bénéfices et que l'État ait à pallier aux coûts qui seront associés aux erreurs et abus engendrés par l'utilisation de l'IA. De nouvelles législations seront aussi nécessaires afin que les gouvernements puissent baliser clairement le développement et l'utilisation de l'IA pour protéger les populations. Les chercheurs, quant à eux, auront la responsabilité de soulever le débat, d'expliquer et de pointer les enjeux, car c'est eux qui détiennent la connaissance. Finalement, les citoyens auront aussi la responsabilité de s'informer et de se questionner sur l'introduction de l'IA dans leur vie quotidienne.

En conclusion, voulons-nous vraiment que l'IA devienne responsable de nous ? N'est-ce pas un oxymore de parler

d'intelligence artificielle responsable ? Est-ce possible de simuler des décisions éthiques ?

IV. Aperçu des interventions du public

Deux périodes de questions ont permis aux membres du public de faire part de leurs interrogations ou réflexions. Pour ce faire, ils devaient utiliser l'adresse courriel soireeia@uqam.ca ou envoyer directement leur intervention sur la plateforme Twitter par l'entremise du mot-clic [#soireeia](https://twitter.com/soireeia). La liste suivante donne un aperçu de ces interventions. Le texte a été modifié pour le rendre anonyme, pour supprimer les mots-clics et pour en augmenter la lisibilité de manière générale.

1) Comment faire en sorte que l'IA ait un impact épistémique positif sur le débat public en évitant les phénomènes de polarisation du débat, les cascades d'information, les biais cognitifs ou les fausses nouvelles ?

2) Un agent artificiel peut-il connaître le bien et le mal ? Dans une société démocratique, peut-on s'en remettre à des IA pour prendre des décisions morales ou éthiques ?

3) Doit-on se préoccuper du pouvoir économique et politique grandissant des grandes entreprises de technologie (ou la Big Tech comme on l'appelle parfois) ? Comment peut-on s'attaquer à ce problème sans affecter l'innovation et la croissance économique ?

4) En quoi les avancées récentes de l'intelligence artificielle couplées à l'avènement des données massives ont-elles un impact sur le respect de la vie privée et les libertés individuelles ?

5) Quel format prendra la formation professionnelle s'il y a personnalisation des cursus d'apprentissage ? Comment peut-on mieux préparer les intervenants de la formation face aux changements à venir ?

6) Quels sont les dangers de l'utilisation de l'IA lors de conflits ? Dans la propagation de discours haineux ? Et dans ce cas comment utiliser l'IA pour contrer ces dangers ?

7) La publicité [#programmatic](https://twitter.com/programmatic) ne mine-t-elle pas déjà les efforts de développement de l'IA responsable ?

8) Comment s'est fait le choix des sept valeurs ?

9) Quelle est la valeur juridique projetée de la Déclaration de Montréal ?

10) C'est quoi, « l'intelligence » ?

11) Le bien-être est culturel, il est différent d'une culture à l'autre : comment l'IA sera en mesure de réconcilier ces différences et les unifier peut-être ?

12) N'y a-t-il pas un paradoxe dans le fait que le créateur de l'IA est faillible et que les attentes soient que l'IA soit infallible ? Veut-on que l'IA apprenne de ses erreurs pour s'améliorer et donc devienne vraiment intelligente ?

13) La suggestion de Hugo Cyr est très pertinente : La Déclaration de Montréal pour l'IA responsable devrait inclure et « promouvoir la justice sociale et l'organisation

collective de la distribution des ressources».

14) Au regard des avancées des technologies de l'information à l'IA, la notion de vie privée ne mérite-t-elle pas d'être repensée ?

15) À des fins décisionnelles, un humain doit-il être obligatoire dans la boucle de l'IA ? Si oui, à quel niveau ?

16) L'humain doit-il apprendre à être plus intelligent que l'IA ? Doit-on brider l'IA de risque que l'humain se fasse déborder par sa création ? N'a-t-on pas juste peur que l'IA devienne meilleure que nous et donc incontrôlable comme dans les films de science-fiction ?

17) Vers la responsabilité algorithmique de notre vie privée : l'IA nous comprend mieux que nous-mêmes. L'être humain peut-il vraiment comprendre le fonctionnement de l'IA et faire des choix éclairés ?

18) Pourquoi l'intelligence humaine et l'intelligence artificielle devraient-elles être similaires ? L'intérêt ne devrait-il pas résider dans la différence ?

19) Si l'IA est un outil, ne devrions-nous pas avoir peur de l'utilisateur de cet outil ? L'ignorance et la crédulité ne sont-elles pas les plus grandes menaces ? Comment nous éduquer et sur quels sujets ? Doit-on apprendre à discriminer le bon du mauvais de l'usage de l'IA ?

20) Comment établir des principes qu'une intelligence artificielle suivra jusqu'au bout dans la mesure où notre perspective est limitée ?

21) Est-ce que la technique du « *differential privacy* » introduite par Apple lors de la WWDC 2016 et une méthode valable pour garantir l'anonymat ? Sinon, quelles sont les grandes options ?

22) Comment concilier l'acceptation sociale qui change si rapidement et des limites (à évolutions lentes) pour déterminer la responsabilité de l'IA ? Et qui les déterminera ? On sait qu'on est déjà sous le joug des grandes entreprises ? Peut-on agir en tant qu'individu ?

23) Est-ce possible de simuler des décisions éthiques ? La responsabilité de l'IA est-elle un oxymore ? Qui est responsable en cas d'erreurs : chercheurs, développeurs, fabricants, consommateurs, ou gouvernements ?

24) Quelle entité pourra contrôler l'IA ?

25) Est-il déjà trop tard pour s'assurer du respect de notre vie privée ? Doit-on miser sur l'éducation des humains pour contrer l'utilisation malveillante de l'IA par d'autres êtres humains ?

26) Comment peut-on garantir que les intérêts d'une IA seront toujours les mêmes que ceux des êtres humains, surtout quand on considère le nouveau statut du robot Sophia qui est devenu citoyen de l'Arabie Saoudite.

27) Quels sont les principes moraux qu'il faudrait mettre en place pour contrer les débordements de l'IA ?

28) Quelles sont les bonnes mesures à prendre pour faire respecter les principes de la Déclaration de Montréal et empêcher des personnes de détourner l'IA à des fins malveillantes ?