

## Vers la régulation de l'intelligence artificielle et la prise en compte des enjeux de discrimination

Regards de l'IEIM | Novembre 2021

Bien que l'intelligence artificielle (IA) soit un sujet dont on entend beaucoup parler, il reste un terme difficile à définir avec précision. Il n'existe en effet pas de réel consensus autour de sa définition. Cela est notamment dû au fait que l'IA a beaucoup évolué et changé de nature depuis sa naissance dans les années 1950<sup>1</sup>. Elle désignait au départ le domaine de recherche qui visait à reproduire artificiellement l'intelligence humaine et à concevoir alors des machines capables de « penser » par elles-mêmes<sup>2</sup>. Cet objectif reste celui de l'IA générale aujourd'hui, mais le terme IA est plus communément utilisé pour désigner l'ensemble des techniques qui permettent d'élaborer des systèmes pouvant effectuer des tâches définies qui sont généralement attribuées aux humains. Les techniques et applications de l'IA étant diverses, il est plus commun pour les chercheurs.e.s de référer directement à une certaine technique ou théorie, comme l'apprentissage automatique par exemple<sup>3</sup>.

*« [...] le terme IA est plus communément utilisé pour désigner l'ensemble des techniques qui permettent d'élaborer des systèmes pouvant effectuer des tâches définies qui sont généralement attribuées aux humains. »*

Malgré les avancées considérables dans beaucoup de domaines que permettent les systèmes d'IA, ces derniers peuvent engendrer des problèmes de discrimination et d'inégalités, notamment lorsqu'ils sont utilisés pour prendre des décisions.

### Un champ d'opportunités... et de risques pour les droits individuels

Les techniques d'IA sont utilisées pour effectuer une grande diversité de tâches. Elles permettent par exemple de recevoir des recommandations personnalisées de films à regarder, d'effectuer des diagnostics médicaux ou encore de produire des contrats ou autres documents légaux. Santé, éducation, administration, marketing, droit, finance, divertissement, sont tant de domaines parmi d'autres qui ont toujours plus recours aux systèmes d'IA, notamment dans le but d'améliorer leurs services. Un grand nombre de décisions prises au quotidien par des compagnies ou des agences gouvernementales reposent aujourd'hui sur des systèmes d'IA.

<sup>1</sup> Le terme « *artificial intelligence* » est apparu pour la première fois lors de la Conférence de Dartmouth aux États-Unis, en 1956, qui réunissait les acteurs-clés du domaine.

<sup>2</sup> C'est à cette conception qu'est liée aujourd'hui l'IA générale (ou « IA forte ») qui désigne des systèmes capables de performer dans toutes les tâches cognitives effectuables par l'humain. Il n'existe pas de tel système de nos jours. Voir « L'IA, c'est quoi ? », *Conseil de l'Europe*, en ligne : <<https://www.coe.int/fr/web/artificial-intelligence/what-is-ai>> (consulté le 15 novembre 2021).

<sup>3</sup> Pour une définition de l'apprentissage automatique, voir par exemple : « Qu'est-ce que l'apprentissage automatique ? Une définition — Apprentissage Automatique », *Data Science* (15 mai 2020), en ligne : <<https://datascience.eu/fr/apprentissage-automatique/question-que-lapprentissage-automatique-une-definition/>> (consulté le 15 novembre 2021).

Ce sont ces systèmes qui vont déterminer, par exemple, quel contenu s'affiche sur les réseaux sociaux pour chacun.e, ou encore, dans le cadre d'une embauche, qui vont recommander les candidatures considérées comme « meilleures » pour un certain poste. Ces systèmes prennent en compte un grand volume d'informations et sont capables de faire des choix (des décisions ou des recommandations) très rapidement sur la base de ces informations.

*« [...] parallèlement à l'enthousiasme engendré par les opportunités de progrès et d'expansion qu'offre l'IA, de nombreuses études ont montré l'existence de risques importants de discrimination inhérents au développement et à l'application des systèmes d'IA ».*

L'idée selon laquelle ces systèmes sont plus précis et plus objectifs que l'activité humaine (subjective, sujette aux erreurs) participe également à la popularité de l'IA pour les prises de décisions. Toutefois, parallèlement à l'enthousiasme engendré par les opportunités de progrès et d'expansion qu'offre l'IA, de nombreuses études ont montré l'existence de risques importants de discrimination inhérents au développement et à l'application des systèmes d'IA.

Un exemple de cela se retrouve dans le contexte des procédures d'embauche où des systèmes d'IA sont souvent utilisés pour recommander les meilleures candidatures pour un certain poste. En 2018, l'agence *Reuters* a publié un article sur un algorithme de ce type développé par Amazon, qui s'est révélé être discriminatoire envers les femmes, en pénalisant leurs candidatures sur la base des embauches passées au sein de la compagnie<sup>4</sup>. Ou encore, dans le secteur de la justice, des algorithmes d'IA sont utilisés pour prédire le comportement de certains individus. Aux États-Unis, une enquête publiée en 2016 a révélé qu'un système d'IA visant à déterminer les risques de récidives des détenus était discriminatoire envers les personnes noires, leur attribuant à tort disproportionnellement plus de risques de récidive qu'aux personnes blanches<sup>5</sup>.

Parfois, la discrimination s'opère de manière plus subtile, et est encore moins facile à identifier. C'est notamment le cas lorsque des algorithmes de recommandation de contenu sont utilisés pour déterminer quel contenu – souvent des annonces publicitaires – sera présenté à quel profil sur les applications ou les sites internet.

L'objectif pour les compagnies qui en font usage est souvent de maximiser leurs revenus. En décidant sur la base d'algorithmes de ne pas présenter certains contenus à certains profils, considérés comme moins à même de consommer lesdits contenus, les compagnies peuvent entraver l'accès à certaines informations qui peuvent s'avérer importantes, notamment des opportunités

<sup>4</sup> Reuters, « Amazon ditched AI recruiting tool that favored men for technical jobs », *The Guardian*, sect. Technology (10 octobre 2018), en ligne : <<https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>> (consulté le 17 juin 2020).

<sup>5</sup> Angwin, J., J. Larson, S. Mattu et L. Kirchner, « Machine Bias », *ProPublica* (23 mai 2016), en ligne : <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=xh2k5je4akpkxjsuxbnhnsjgpcskwu4v>> (consulté le 14 mai 2021).

d'emploi ou de logement<sup>6</sup>. Les décisions prises sur la base des systèmes d'IA comportent des risques importants d'affecter négativement la vie des individus de manière conséquente, et ces exemples l'illustrent bien.

Par ailleurs, des enjeux de transparence et d'explicabilité se mêlent aux risques de discrimination. Les techniques d'IA sont complexes et leur caractère automatique ne permet pas toujours de saisir les « raisonnements » qui ont conduit à une décision ou à une recommandation ; c'est le principe de la « boîte noire ». Deux implications principales en découlent. Premièrement, il n'est pas toujours évident d'identifier les cas de discrimination. Dans les exemples présentés ci-dessus, des études ont été menées expressément pour savoir si les algorithmes étaient discriminatoires, mais il est concevable qu'une personne victime d'une décision discriminatoire ne s'en rende pas compte<sup>7</sup>. Deuxièmement, même lorsque des traitements discriminatoires sont identifiés, il est parfois difficile de comprendre comment un système d'IA est parvenu à un tel résultat, et ce même pour les ingénier.e.s ayant codé l'algorithme. Il devient donc difficile pour une personne de prouver qu'elle a été discriminée dans le cadre d'une décision la concernant, et donc difficile d'obtenir réparation. Tout cela engendre une forme d'automatisation des biais et des systèmes discriminatoires, ce qui, bien entendu, va à l'encontre de la lutte contre les discriminations<sup>8</sup>.

### L'éthique et le droit pour mitiger les risques de discrimination liés à l'IA

Dans ce contexte, une multitude d'initiatives ont vu le jour autour du globe, afin d'assurer un développement et une utilisation responsable de l'IA. Elles se présentent principalement comme des cadres éthiques - il en existe plus d'une centaine à ce jour – sous la forme de déclarations, de livres blancs, de codes de conduite ou encore de chartes, à l'instar de la Déclaration de Montréal pour un développement responsable de l'IA lancée en 2017<sup>9</sup>. Bien que ces initiatives aient permis de mettre en lumière une multitude d'enjeux de plus en plus pressants liés à l'IA, dont les risques de discrimination, et d'offrir des terrains de réflexion et de discussion sur de possibles réponses aux enjeux identifiés, elles sont de nature non-contraignante et présentent souvent des principes abstraits et généraux qui sont difficiles à traduire en termes techniques et donc à intégrer dans la conception et l'exploitation des systèmes d'IA.

Dans le contexte de la multiplication de ces cadres éthiques, les lacunes du droit face aux risques concrets de l'IA ont été davantage mises en évidence. Malgré le caractère fondamental du droit

---

<sup>6</sup> Aux États-Unis, une décision administrative avait été rendue contre Facebook au sujet du fonctionnement du systèmes d'annonces pour des logements, voir notamment Tracy JAN et Elizabeth DWOSKIN, « HUD is reviewing Twitter's and Google's ad practices as part of housing discrimination probe », *Washington Post*, en ligne : <<https://www.washingtonpost.com/business/2019/03/28/hud-charges-facebook-with-housing-discrimination/>> (consulté le 10 novembre 2021). Un cas similaire avait été engagé au Canada au sujet d'annonces pour des emplois. Voir notamment : Elizabeth Thompson · CBC NEWS, « Facebook faces prospect of class action over job ads | CBC News », *CBC* (11 avril 2019), en ligne : <<https://www.cbc.ca/news/politics/facebook-job-ads-class-action-1.5094661>> (consulté le 10 novembre 2021). Sur le sujet, voir aussi : Olivier SYLVAIN, « Opinion | A Watchful Eye on Facebook's Advertising Practices », *The New York Times*, sect. Opinion (28 mars 2019), en ligne : <<https://www.nytimes.com/2019/03/28/opinion/facebook-ad-discrimination-race.html>> (consulté le 10 novembre 2021); Jinyan ZANG, « Solving the problem of racially discriminatory advertising on Facebook », *Brookings* (19 octobre 2021), en ligne : <<https://www.brookings.edu/research/solving-the-problem-of-racially-discriminatory-advertising-on-facebook/>> (consulté le 10 novembre 2021).

<sup>7</sup> Cela s'avère particulièrement problématique pour les systèmes d'IA qui réintègrent les résultats produits (les *outcomes*) comme données d'entraînement (les *inputs*) pour effectuer une certaine tâche.

<sup>8</sup> Voir Virginia EUBANKS, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Publishing Group, 2018; Cathy O'NEIL, *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*, New York, Crown, 2016.

<sup>9</sup> <https://www.declarationmontreal-iaresponsable.com/>

à l'égalité et à la non-discrimination au sein d'une majorité de systèmes, il est apparu que le droit en place ne permettait pas de saisir la nature complexe des enjeux liés à l'IA, notamment en ce qui a trait aux nouvelles formes de discrimination que cette dernière entraîne. Face à cela, quelques États se sont emparés de la question et ont - plus ou moins timidement - commencé à légiférer.

En 2016 est entré en vigueur le Règlement général sur la protection des données – le fameux RGPD- de l'Union européenne<sup>10</sup>. Ce texte ne porte pas directement sur l'IA et ses enjeux, mais il s'intéresse aux données, éléments très importants en IA. Plus précisément, il vise le traitement des données à caractère personnel dans un objectif de protection des droits individuels, comme le droit à la vie privée. Il vise également certaines pratiques de profilage algorithmique<sup>11</sup> qui participent à la discrimination. Ce règlement qui prévoit d'importantes sanctions financières en cas de manquements, a inspiré plusieurs États à légiférer en la matière, dont le Canada avec le projet de loi C-11, et le Québec avec le projet de loi n°64 sur la protection des renseignements personnels. Toutefois, malgré les avancées que constituent ces textes en termes de protection des droits, ils ne reflètent pas les réalités de la pratique en termes de fonctionnement et d'application des systèmes d'IA et présentent donc quelques limites au regard de la prise en charge des risques de discrimination.

*« Cette directive [gouvernementale canadienne] exige notamment une évaluation algorithmique des systèmes d'IA préalable à leur utilisation [...] »*

Semblant être conscient de la nécessité de traiter plus concrètement de l'IA, le Canada s'est précisément penché sur la question de la prise de décision automatisée. Ainsi, en 2019 est entrée en vigueur une directive encadrant cette pratique au sein des agences gouvernementales. Cette directive exige notamment une évaluation algorithmique des systèmes d'IA préalable à leur utilisation. Elle impose également des exigences de transparence et de contrôle qui visent à « veiller à ce que les systèmes décisionnels automatisés soient déployés d'une manière qui permet de réduire les risques pour les Canadiens et les institutions fédérales, et qui donne lieu à une prise de décision plus efficace, exacte et conforme »<sup>12</sup>. Toutefois, ces exigences ne s'imposent qu'aux agences fédérales et ne concernent pas les compagnies privées qui pourtant font un usage important des systèmes d'IA d'aide à la décision.

Plus récemment, la Commission européenne a présenté une proposition de texte juridique portant plus largement sur l'IA. Il s'agit de la *Législation sur l'IA* de la Commission européenne qui a pour objectif principal « que les systèmes d'IA utilisés dans l'Union européenne soient sûrs, transparents, éthiques, impartiaux et sous contrôle humain »<sup>13</sup>, c'est une première dans son genre.

<sup>10</sup> « Protection des données dans l'UE », *Commission européenne*, en ligne : <[https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en)> (consulté le 15 novembre 2021).

<sup>11</sup> Le profilage algorithmique consiste à utiliser des données pour évaluer certains aspects personnels relatifs à une personne, notamment pour analyser ou prédire des éléments concernant sa situation économique, sa santé, ses préférences personnelles, ses intérêts, son rendement au travail, ses fiabilités, ses déplacements et plus généralement ses comportements. Des décisions sont souvent prises sur la base de ces analyses ou prédictions.

<sup>12</sup> SECRÉTARIAT DU CONSEIL DU TRÉSOR DU CANADA, *Directive sur la prise de décisions automatisée*, 2019, paragraphe 4.1.

<sup>13</sup> « Excellence et confiance en matière d'intelligence artificielle », *Commission européenne*, en ligne : <[https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence\\_fr](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_fr)> (consulté le 10 novembre 2021).

Pour ce faire, le texte prévoit une classification des systèmes d'IA en fonction de leur niveau de risque, et des règles spécifiques pour chacune des catégories de risques : inacceptable, haut, limité et minimal. Les systèmes d'IA dont le niveau de risques est jugé inacceptable sont tout simplement interdits. Les systèmes d'IA utilisés pour noter les citoyen.ne.s afin d'établir un classement de fiabilité selon lequel des traitements différentiels seront opérés sont par exemple visés par cette interdiction. Ceux dont le niveau de risque est « haut » sont autorisés, mais à condition de respecter des exigences de contrôle et de transparence. Ils doivent par exemple faire l'objet de procédures d'évaluation avant leur mise en marché, et il y a une obligation pour les fournisseurs d'informer le public lorsqu'il interagit avec un système automatisé. Il y a peu d'exigences pour les systèmes présentant des niveaux de risque « limité » ou « minimal » mais la législation encourage les fournisseurs de ces systèmes à créer des codes de conduite sur la base des exigences qui imputent aux systèmes d'IA à haut risque. La Commission européenne s'applique par ailleurs à bien définir les termes et concepts clés de l'éthique de l'IA, ce qui est particulièrement pertinent dans un domaine aussi dynamique et vaste que l'IA.

Les organisations internationales se sont également emparées de la question et ont publié divers textes qui visent à assurer un développement et une utilisation responsable de l'IA, en prenant en compte les risques de discrimination. La dernière en date est un projet de recommandation de l'UNESCO sur l'éthique de l'IA<sup>14</sup>. L'organisation la qualifie comme « le premier instrument normatif mondial sur l'éthique de l'IA ». Même si les recommandations ne sont pas contraignantes pour les Etats membres, elles sont empreintes d'un fort impact doctrinal qui pourrait guider les États pour légiférer sur le sujet.

Ces différents textes semblent traduire un réel intérêt de la part des États pour la problématique, et une intention de mitiger les risques, notamment à l'aide d'instruments juridiques. Toutefois, ces derniers sont encore en phase de développement et d'implantation alors que les techniques d'IA et leurs champs d'application se développent à un rythme effréné. De plus, même s'il semble y avoir un consensus sur la nécessité de considérer les implications de l'IA, une légère cacophonie semble marquer ces différentes initiatives : tantôt elles se concentrent sur les besoins de transparence, tantôt sur les données, sur la défense des droits individuels ou sur le soutien à l'innovation... Complexes, fluctuants et transversaux, il est extrêmement difficile d'adresser juridiquement tous les enjeux liés à l'IA.

*« Ces derniers [textes juridiques] sont encore en phase de développement et d'implantation alors que les techniques d'IA et leurs champs d'application se développent à un rythme effréné. [...] »*

### **Des risques concrets qui requièrent des réponses concrètes et effectives**

Ce rapide tour d'horizon des enjeux de discrimination liés à l'IA et de la régulation de ce domaine permet de saisir plusieurs éléments principaux sur les enjeux discriminatoires de l'IA.

<sup>14</sup> Le projet de recommandation va être soumis à la Conférence générale de l'UNESCO 2021. Voir « Élaboration d'une Recommandation sur l'éthique de l'intelligence artificielle », UNESCO (27 février 2020), en ligne : <<https://fr.unesco.org/artificial-intelligence/ethics>> (consulté le 10 novembre 2021).

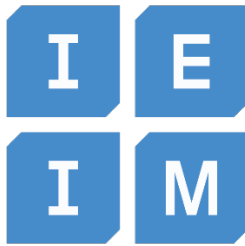
- Tout d'abord, que de vrais risques de discrimination et d'exacerbation des inégalités existent et qu'il est nécessaire de les adresser comme tels, et non comme des dommages potentiels. Il est important de cadrer concrètement les pratiques actuelles et du futur proche.
- Ensuite, ce tour d'horizon permet de comprendre la complémentarité du droit et de l'éthique, et le besoin de collaboration entre divers secteurs et disciplines.
- L'éthique se veut moins contraignante que le droit, mais elle permet de poser les bornes de ce qui est souhaitable ou non pour la société. Elle guide également dans des situations où le droit est silencieux ou non applicable<sup>15</sup>.
- Le droit, lui, doit s'adapter quand surviennent de nouvelles formes de discrimination, ce qui est un défi de taille dans le contexte de l'IA qui évolue si rapidement. Il faut également garder en tête qu'il ne s'agit pas ici d'enjeux purement techniques mais qu'un regard interdisciplinaire est nécessaire. Les discussions sur les implications de l'IA font entre autres appel à des expert.e.s en droit, en sociologie, en informatique, en économie ou encore en psychologie.
- Pour finir, il est important de considérer qu'au regard de la diversité et de la complexité des enjeux de discrimination liés à l'IA, il n'est pas possible de considérer une réponse unique. L'élaboration de plusieurs cadres normatifs complémentaires est nécessaire.

---

<sup>15</sup> Par exemple, dans le cadre de la législation sur l'IA de la Commission européenne, il n'y a pas de règles strictes pour les systèmes présentant un risque minimal (et qui sont les plus répandus aujourd'hui), mais les fournisseurs et exploitants de ces systèmes sont encouragés à établir des règles de conduite par eux-mêmes.

Institut d'études internationales de Montréal  
Université du Québec à Montréal  
400, rue Sainte-Catherine Est  
Bureau A-1540, Pavillon Hubert-Aquin  
Montréal (Québec) H2L 3C5  
514 987-3667  
[ieim@uqam.ca](mailto:ieim@uqam.ca)  
[www.ieim.uqam.ca](http://www.ieim.uqam.ca)

**UQÀM**



**Institut d'études  
internationales  
de Montréal**

Autrice

**Alexandra Parada**

**Doctorante en droit, Département des sciences juridiques de l'UQAM**