

Université de Montréal

**La performativité algorithmique :
Construction identitaire et interférence**

Par :

Alexis Beaupré-Daignault

Département de Philosophie
Faculté des Arts et Sciences

Mémoire présenté en vue de l'obtention du grade de maîtrise en philosophie
Option philosophie au collégial

Août 2022

©Alexis Beaupré-Daignault, 2022

Ce mémoire intitulé

La performativité algorithmique : Construction identitaire et interférence

Présenté par

Alexis Beaupré-Daignault

A été évalué par un jury composé des personnes suivantes

Molly Kao

Présidente du jury

Jonathan Simon

Membre du jury

Marc-Antoine Dilhac

Directeur de recherche

Résumé

Dans ce mémoire, nous tentons de déterminer si les SIA participent à la détérioration de la justice sociale. En ce sens, notre hypothèse est qu'une force performative émerge de la répétition des décisions algorithmiques, laquelle interfère avec la construction identitaire des individus. Selon le cadre d'analyse préconisé, lequel s'inscrit au sein de la conception de la justice d'Iris Marion Young, nous soutenons que l'interférence identitaire est injuste puisqu'elle contribue à l'oppression de l'impérialisme culturel. S'avérant, ce phénomène de performativité algorithmique expliquerait, d'une part, l'interaction entre l'identité et les algorithmes, et permettrait d'autre part son exploitation. Ultimement, nous soutenons que la performativité des algorithmes pourrait être manipulée afin qu'elle soit mise au service de la justice sociale.

Mots clés : algorithme — biais algorithmique — éthique de l'IA — identité — justice sociale — neutralité — performativité — représentation

Abstract

In this thesis, we try to determine if artificial intelligence systems contribute to the deterioration of social justice. In this sense, our hypothesis is that a performative force emerges from the repetition of algorithmic decisions, which interferes with the identity construction of individuals. According to the recommended analytical framework, which is part of Iris Marion Young's conception of justice, we argue that identity interference is unjust because it contributes to the oppression of cultural imperialism. If it turns out, this phenomenon of algorithmic performativity would explain, on the one hand, the interaction between identity and algorithms, and on the other hand allow the exploitation of the latter. Ultimately, we argue that the performativity of algorithms could be manipulated so that it is put at the service of social justice.

Key words: algorithm — algorithmic bias — identity — IA ethics — social justice — neutrality — performativity — representation

Table des matières

Introduction	8
1 Développement de l'IA et implications.....	13
1.1 Les données	13
1.1.1 Le « Big Data ».....	14
1.1.2 L'internet des objets	15
1.1.3 Les données personnelles	16
1.1.4 L'anonymisation/pseudonymisation des données	18
1.2 Les algorithmes et le critère d'intelligibilité.....	19
1.2.1 Les algorithmes déterministes	20
1.2.2 Les algorithmes apprenants	21
1.2.3 L'apprentissage machine	21
1.3 L'intelligence artificielle aujourd'hui.....	23
1.3.1 L'apprentissage profond.....	23
1.3.2 L'analogie des boîtes noires	25
1.3.3 La compréhension globale des effets.....	27
2 Performativité algorithmique et construction identitaire.....	29
2.1 La performativité.....	29
2.1.1 L'origine du concept.....	29
2.1.2 La performativité réinterprétée	31
2.2 La construction performative de l'identité	37
2.2.1 Les groupes sociaux et l'identité performative.....	37
2.2.2 L'intersectionnalité.....	39
2.2.3 L'identité numérique	42
2.3 La performativité algorithmique.....	48
2.3.1 Le concept d'émergence.....	49
2.3.2 Les complexes algorithmiques	50
2.3.3 Le renforcement performatif	53
2.3.4 La socialisation de l'identité.....	55
3 Normes de justice et biais algorithmiques.....	60
3.1 L'injustice algorithmique	60
3.2 Les conceptions de justice	62
3.2.1 La « <i>Théorie de la Justice</i> ».....	62
3.2.2 La Justice et la « <i>politique de la différence</i> »	68
3.3 L'interférence identitaire	77

3.3.1	Le mythe de la neutralité algorithmique	78
3.3.2	Les biais algorithmiques	79
3.3.3	Les boucles de rétroaction	83
3.3.4	La stigmatisation et la conformité	85
4	Pour une construction identitaire sans interférence	87
4.1	L'échelle d'analyse.....	88
4.1.1	Le dynamisme de l'IA	89
4.2	L'affirmation de la différence.....	91
4.2.1	La différence de traitement.....	91
4.2.2	Le rejet de l'impartialité	92
4.2.3	La redéfinition de l'égalité	93
4.2.4	Les politiques fondées sur la reconnaissance des groupes	94
4.3	L'inclusivité et la représentativité	95
4.3.1	Les enjeux de représentativité au sein de l'IA.....	96
4.3.2	L'éducation et la sensibilisation	97
4.3.3	La diversité et la discrimination positive.....	98
4.4	La Coalition Arc-en-ciel Représentative de l'Intelligence Artificielle (CARIA).....	99
4.4.1	La classification des complexes algorithmiques.....	100
4.4.2	Les méthodes d'exploitation de la performativité algorithmique.....	102
	Conclusion.....	106
	Bibliographie	111

Remerciements

À Marc-Antoine Dilhac pour avoir cru en moi ainsi qu'en ce projet.

À ma mère pour son soutien généreux.

À ma grand-mère pour sa relecture minutieuse et sa bienveillance.

À ma sœur pour son amour immuable.

À ma partenaire de vie sans qui je n'aurais jamais trouvé le courage ni détermination de mener à bien cette entreprise.

À cette deuxième famille m'ayant accueilli comme un fils.

Introduction

« Le coupe-papier est à la fois un objet qui se produit d'une certaine manière et qui, d'autre part, a une utilité définie; et on ne peut pas supposer un homme qui produirait un coupe-papier sans savoir à quoi l'objet va servir [...] Nous dirons donc que, pour le coupe-papier, l'essence [...] précède l'existence »¹. Le développement de l'intelligence artificielle (IA) est paradoxal puisque l'intention derrière sa production est nouvellement indéfinie étant donné ses applications potentiellement infinies. Même si l'être humain est à la base de sa création et de sa programmation, l'IA se distingue des autres outils étant donné ses capacités d'automatisation et d'apprentissage. C'est donc un objet dont l'utilité est virtuellement infinie, et dont l'essence ne précède pas l'existence. Nous croyons toutefois qu'il est faux d'affirmer que son existence précède son essence puisque cela lui conférerait un statut similaire à celui de l'homme, celui d'un être libre². Malgré le statut singulier de cette technologie, l'autonomie et le potentiel évolutif de l'IA sont ultimement limités par l'agentivité humaine. Conséquemment à mi-chemin entre l'être humain et l'outil, l'IA soulève nombre d'enjeux moraux et éthiques étant donné l'étendue de ses utilisations possibles.

Étant avant tout une création humaine, l'IA est destinée à occuper la place qui lui sera socialement accordée, et c'est pour cette raison que la question de la réglementation et de l'encadrement du déploiement de l'IA est d'actualité. S'il est envisageable de laisser aller, sans aucune restriction, le déploiement de l'IA, il est également possible de défendre que son abolition est souhaitable. Entre ces deux positions aux antipodes s'étend un spectre de possibilités au sein duquel une position intermédiaire est d'ailleurs présentée par *La Déclaration de Montréal pour un Développement responsable de l'Intelligence Artificielle*³ publiée en 2018. Il s'agit d'une initiative entreprise par l'Université de Montréal ayant pour objectif de mettre de l'avant des principes⁴ dont le respect permettrait

¹ SARTRE, Jean-Paul, *L'existentialisme est un humanisme*, Paris, 1946, p. 26-27.

² *Ibid.*

³ Déclaration de Montréal IA responsable, *La déclaration de Montréal pour un Développement responsable de l'Intelligence Artificielle*, 2018.

⁴ Dix principes éthiques sont répertoriés dans *La Déclaration de Montréal*, soit ceux de la prudence, du respect de l'autonomie, de la protection de l'intimité et de la vie privée, de la solidarité, de la participation démocratique, de l'équité, de l'inclusion de la diversité, de la prudence, de la responsabilité, et du développement soutenable.

le développement éthique et responsable de l'IA. Ce mémoire se positionne dans le prolongement idéologique de cette déclaration puisque nous jugeons qu'en dépit des risques inhérents à l'implantation de ces nouvelles technologies, l'IA constitue avant tout un outil dont l'utilité potentielle pour l'humanité est un gage des risques qu'il représente pour la justice sociale. Concrètement, le rôle de *La Déclaration de Montréal* est ultimement de minimiser ces risques en délimitant les contours du développement éthique et responsable de l'IA. Grâce à une perspective hybride entre le conséquentialisme et le déontologisme, l'objectif est de s'assurer que l'action des algorithmes génère plus de bienfaits que de conséquences négatives pour les êtres humains, et qu'elle respecte certaines règles absolues. À titre d'exemple, il y est indiqué que « les SIA [systèmes d'intelligence artificielle] ne doivent pas constituer une source de mal-être, sauf si ce dernier permet d'engendrer un bien-être supérieur que l'on ne peut atteindre autrement »⁵. Cette règle implique nécessairement un calcul des conséquences, tandis qu'il est également possible d'y trouver des règles éthiques, sous la forme d'impératifs catégoriques, comme c'est le cas de celle indiquant que « la décision de tuer doit toujours être prise par des êtres humains et la responsabilité de cette décision ne peut être transférée à un SIA »⁶. Si nous jugeons que cette approche éthique hybride est cohérente avec les exigences que nous devrions avoir face au déploiement de l'IA, nous soutenons néanmoins que notre compréhension de ses effets sur la justice sociale est insuffisante afin d'en comparer quantitativement les répercussions positives et négatives.

Afin de nous intéresser aux implications du développement de l'IA sur la justice sociale, nous nous poserons la question suivante : *En quoi l'utilisation grandissante des processus décisionnels automatisés est-elle susceptible d'engendrer un accroissement des injustices structurelles?* Si la littérature répertoriant les cas d'injustices algorithmiques est importante et relativement complète, il faut néanmoins souligner que leur analyse n'est généralement pas située au sein de cadres de justice permettant leur compréhension théorique. Malgré cette tendance que nous avons pour objectif de renverser, il est clair que l'intuition à la base de ce mémoire est devenue possible grâce aux pionniers qui ont su déceler ces injustices nouvelles qui s'immiscent au sein de la société; je pense notamment

⁵ *Ibid.*, p. 8.

⁶ *Ibid.*, p. 16.

à Safiya Noble qui explique comment les moteurs de recherches participent au racisme et au sexisme dans *Algorithms of Oppression : How Search Engines Reinforce Racism* publié en 2018, à Andrew Ferguson qui fait la lumière sur l'automatisation de la discrimination policière dans *The Rise of Big Data Policing : Surveillance, Race, and the Future of Law Enforcement* publié en 2017, ainsi qu'à Virginia Eubanks qui explique comment la gestion technologique de la pauvreté impacte les groupes économiquement désavantagés dans *Automating Inequality, How High-Tech tools profile, police, and punish the poor* publié en 2018. En ce sens, s'il est déjà démontré que les décisions algorithmiques génèrent des situations d'injustices, nous estimons de surcroît que ces dernières s'accumulent pour impacter la justice sociale. Afin de soutenir cela, notre hypothèse est qu'une force performative contraignante émerge de l'accumulation des décisions algorithmiques, phénomène que nous qualifierons de performativité algorithmique et qui s'inspire de la notion de performativité réinterprétée par Judith Butler en 1990 dans *Gender Trouble : Feminism and the Subversion of Identity*. Si nous sommes d'avis que *l'utilisation grandissante des processus décisionnels automatisés engendre un accroissement des injustices structurelles*, c'est donc parce que nous présumons que la performativité algorithmique interfère avec la construction identitaire des individus. Grâce à la notion d'impérialisme culturel tel que théorisée par Iris Marion Young en 1990 dans *Justice and the Politics of Difference*, nous soutiendrons que cette interférence identitaire participe à l'exacerbation des injustices structurelles en favorisant certains contenus représentationnels au détriment des autres qu'elle marginalise. Si les effets contraignants et injustes du phénomène de performativité algorithmique s'expliquent par sa contribution au renforcement de l'impérialisme culturel, nous persistons à croire qu'il est possible d'exploiter la force performative des algorithmes afin qu'elle contribue inversement à l'atteinte d'une plus grande justice sociale.

En prévision d'une telle démonstration, nous allons, au premier chapitre, décrire le fonctionnement de l'IA grâce à l'explicitation de ses composantes. L'objectif sera d'être en mesure d'établir que les processus décisionnels automatisés engendrent des répercussions réelles, et potentiellement nuisibles, chez les individus. Nous poursuivrons en démontrant que l'inintelligibilité des algorithmes, s'expliquant notamment grâce à la complexification de l'apprentissage-machine et de l'apprentissage profond, confère un

statut particulier au processus décisionnels automatisés, celui d'*eventful* ou d'accomplissement en devenir. Cette particularité des algorithmes sera effectivement centrale pour les considérations qui suivront. Dans le deuxième chapitre, nous démontrerons que le phénomène de la performativité algorithmique nous permet d'expliquer que l'action des processus décisionnels automatisés a une incidence directe sur la construction identitaire des individus. Nous défendons donc qu'une force performative émerge de l'orchestration du nombre grandissant de décisions algorithmiques continuellement générées, laquelle aurait ultimement un effet contraignant sur la construction identitaire des individus en les classifiant selon des catégories générées artificiellement. Au troisième chapitre, notre objectif sera d'établir que le phénomène de la performativité algorithmique participe à la détérioration de la justice sociale. Notre première tentative en ce sens sera d'examiner la *Théorie de la Justice* rawlsienne afin de déterminer si elle rend possible le traitement des enjeux soulevés par le développement de l'IA. Une fois contraint à l'insuffisance théorique du paradigme distributif, nous considérerons l'approche critique d'Iris Marion Young qui nous permettra, notamment grâce à la théorisation des concepts groupes sociaux et d'oppression, de défendre que la performativité algorithmique constitue une injustice puisqu'elle interfère avec la construction identitaire. Lors du quatrième chapitre, nous montrerons inversement que le phénomène de la performativité algorithmique n'est pas fondamentalement injuste, mais qu'il devient plutôt contraignant lorsqu'il participe au renforcement de l'impérialisme culturel. À l'image du concept de discrimination mobilisée par Young, nous soutenons qu'il est possible d'orienter artificiellement les biais algorithmiques de manière à ce que le pouvoir performatif des algorithmes contribue à l'émancipation des membres de groupes stigmatisés. Afin de mettre la performativité algorithmique au service de la justice sociale, nous proposerons la mise sur pied d'une institution qui aura pour objectif de s'assurer que les répercussions sociales des algorithmiques soient minimalement positives et désirables; nous l'appellerons Coalition arc-en-ciel Représentative de l'Intelligence Artificielle (CARIA).

En ce sens, nous soutenons non seulement que le déploiement responsable de l'IA est théoriquement possible, mais nous croyons qu'il est également possible d'orienter son développement afin qu'il contribue à l'atteinte d'une plus grande justice sociale. En

postulant que l'IA est un outil dont l'opérationnalisation est dépendante de l'intentionnalité humaine, nous soutenons effectivement que l'être humain détient les moyens pour s'assurer que la technologie lui soit collectivement et socialement profitable. Si le progrès technologique est historiquement apparu comme une arme à double tranchant, nous croyons qu'il est du devoir de la communauté scientifique d'en anticiper les dérives potentiellement néfastes afin de minimiser leurs effets. Assurément, la délimitation entre les répercussions positives et négatives du déploiement de l'IA dépend ultimement des conceptions du bien et/ou du juste qu'une société poursuit, et c'est justement pour cette raison qu'une interprétation philosophique de cet enjeu est nécessaire si l'on souhaite pouvoir départager ce qui est désirable d'indésirable d'un point de vue sociétal. Bien que nous soyons pertinemment conscients des bienfaits que peut manifestement générer le projet de l'IA, nous allons nous concentrer sur ses effets indésirables afin d'éclaircir la frontière les délimitant. À partir du moment où les SIA dont les décisions sont problématiques sont isolés, il devient possible de soutenir que les autres sont acceptables puisqu'elles ne participent pas à la détérioration de la justice sociale. À ce titre, nous nous focaliserons sur une norme de justice particulière, celle de la libre construction identitaire, afin de déterminer si l'usage d'un SIA est justifiable socialement. Cette focalisation nous permettra effectivement d'avoir une vue d'ensemble des répercussions négatives des SIA sur la société.

1 Développement de l'IA et implications

Avant de pouvoir démontrer que le traitement de masse des données personnelles est susceptible d'engendrer un accroissement des injustices structurelles, il est nécessaire d'expliquer comment le projet de l'IA a concrètement des implications sur la vie des gens. En 1950, Alan Turing fait déjà référence à *l'intelligence des machines* en développant un test ayant pour objectif de déterminer si les machines pensent.⁷ La puissance des ordinateurs s'étant exponentiellement amplifiée au cours des soixante-dix dernières années, les machines *intelligentes* font aujourd'hui partie intégrante de nos vies. Le concept de l'IA est néanmoins difficile à définir étant donné qu'il fait indirectement référence à l'intelligence, représentation qui, encore aujourd'hui, comporte son lot de complications. Tout comme l'intelligence humaine, l'explication de l'IA nécessite la clarification de plusieurs composantes dont l'articulation en illustre le fonctionnement. À cette fin, nous utiliserons la définition qui suit : « L'intelligence artificielle (IA) désigne l'ensemble des techniques qui permettent à une machine de simuler l'intelligence humaine, notamment pour apprendre, prédire, prendre des décisions et percevoir le monde environnant. Dans le cas d'un système informatique, l'intelligence artificielle est appliquée à des données numériques. »⁸ À l'image de la perception sensorielle, l'IA nécessite un véhicule lui permettant d'exister au sein de la réalité matérielle et/ou virtuelle. Les systèmes d'intelligence artificielle (SIA) représentent donc « tout système informatique utilisant des algorithmes d'intelligence artificielle, que ce soit un logiciel, un objet connecté ou un robot »⁹, permettant à l'IA d'avoir une incidence décisionnelle sur la réalité.

1.1 Les données

L'une des caractéristiques inhérentes de la quatrième révolution industrielle¹⁰ fait référence à l'essence immatérielle de son combustible virtuel, la donnée. La métaphore du carburant est assurément évocatrice afin de définir la donnée puisqu'elle permet de saisir explicitement son utilisation. Effectivement, tout comme la valeur du pétrole découle directement de son utilisation rendue possible grâce au moteur à combustion, la valeur

⁷ TURING, Alan, *Computing Machinery and Intelligence*, 1950, p. 433-460.

⁸ Déclaration de Montréal IA responsable, *La déclaration de Montréal pour un Développement responsable de l'Intelligence Artificielle*, 2018, Lexique.

⁹ *Ibid.*

¹⁰ SENDLER, Ulrich, *The internet of things: Industrie 4.0 Unleashed*, 2018, p. 9.

nouvellement attribuée à la donnée est le fruit de la puissance des ordinateurs permettant son exploitation. Néanmoins, ces deux types de ressources sont intrinsèquement différentes, car l'une existe à l'état naturel tandis que l'autre est, en toute fin pratique, une création virtuelle. La donnée existe essentiellement grâce à deux processus, celui de sa création faisant référence à l'ensemble des capteurs permettant de saisir l'information, et celui de son stockage représentant la capacité d'enregistrer les informations dans un objectif d'exploitation. Les systèmes d'acquisition et d'archivage des données (SAAD) « désignent tout système informatique pouvant collecter et enregistrer des données. Celles-ci seront éventuellement utilisées pour l'entraînement d'un système d'IA ou comme paramètres pour une prise de décision. »¹¹ Par définition, une donnée est donc « une information codée, figée et transmissible »¹² alors que la notion de donnée numérique, à laquelle nous ferons référence pour l'ensemble de nos considérations, renvoie spécifiquement à « la donnée dont la mémorisation a été confiée à un système informatique (et non à un papier, par exemple). »¹³ Concrètement, les données constituent donc des informations touchant n'importe quelle facette du monde réel ou virtuel dont l'acquisition et l'archivage est exécutée par un SAAD.

1.1.1 Le « Big Data »

L'archivage des données, ou informations, n'est pas propre à l'ère numérique puisque les livres ont longtemps assuré cette fonction. À l'évidence, les méthodes d'archivage traditionnelles sont coûteuses et inefficaces lorsqu'on les compare avec les moyens technologiques de la révolution numérique actuelle.¹⁴ C'est en réduisant considérablement les contraintes économiques qu'impliquait le stockage de l'information que les technologies numériques ont rendu possible la conservation massive des données.¹⁵ Face à l'essor du *Big Data* (exploitation de très grands volumes de données¹⁶) émanant de

¹¹ *Ibid.*

¹² PERES, Éric, *Les données numériques, un enjeu d'éducation et de citoyenneté*, Paris, Les éditions des journaux officiels, 2015, p .10.

¹³ *Ibid.*

¹⁴ La bibliothèque du Congrès des États-Unis considérée comme la plus importante au monde rassemble 23 millions d'ouvrages. L'équivalent numérique des ouvrages stockés dans cette bibliothèque pourrait aujourd'hui être stockée sur 20 terabits (soit l'équivalent de quelques disques durs de haute capacité). JOHNSTON, Leslie, «A "Library of Congress" Worth of Data: It's All In How You Define It», *Library of Congress*, 2012.

¹⁵ PERES, Éric. Op. cit., p .6.

¹⁶ *Ibid.*, p .9.

ces nouvelles conditions de possibilités, la nature des informations archivées s'étend encore à ce jour. Effectivement, les données numériques correspondent désormais à l'entièreté des phénomènes observables, allant des comportements et habitudes humaines aux manifestations et événements naturels. Le *data* représente donc toutes les informations numériques ou alphanumériques codées et uniquement lisibles par la machine pour son enregistrement, son traitement, sa conservation ou encore sa communication¹⁷. Quant à elle, « la *datafication* est plus que la simple mesure ou transformation du réel en *datas* qui peuvent être recueillies et analysées, c'est la perception qu'il y a, dans les grands nombres de données, plus de valeurs que dans les petits volumes de *datas*. »¹⁸ À titre indicatif, IBM avançait, en 2012, que 90% de l'entièreté des données collectées et archivées dans le monde avaient été créées au cours des deux dernières années¹⁹. Certes, cette accumulation exponentielle des données numériques est justifiée, non seulement, grâce au faible coût de nos moyens technologiques, mais également, à cause de la valeur des actions qu'elle rend possibles. Grâce au traitement des algorithmes qui sera décrit plus loin, « le Big Data permet de relever des corrélations statistiques régulières et récurrentes entre des phénomènes »²⁰ lesquelles se traduisent ultimement par une faculté de prédiction probabiliste²¹ qui est, pour sa part, monnayable.

1.1.2 L'internet des objets

Initialement, les données numériques récoltées étaient associées au déploiement de l'internet puisqu'elles faisaient uniquement référence aux informations provenant de la réalité virtuelle, laquelle se limitait aux actions posées sur les réseaux en ligne. Seuls les

¹⁷ *Ibid.*

¹⁸ *Ibid.*, p .10.

¹⁹ KLEYMAN, Bill, « The big Data Battleground: Analyzing the Big Picture », *DataCenter Knowledge*, mis à jour le 12 septembre 2012, <https://www.datacenterknowledge.com/archives/2012/09/12/a-look-into-the-big-data-battleground-analyzing-the-market?utm-source=feedburner&utm-medium=feed&utm-campaign=Feed%3A+DataCenterKnowledge+%28Data+Center+Knowledge%29>, consulté le 8 octobre 2020.

²⁰ PERES, Éric. *Op. cit.*, p .56.

²¹ « Le Big Data permet de relever des corrélations statistiques et donc de faire une prédiction probabiliste. Concrètement, avec le Big Data, on ne sait pas toujours pourquoi telle chose va se produire, mais on peut la prévoir avec une forte probabilité. Par exemple, on peut, en temps réel, et avec une géographie précise, savoir où se trouvent les personnes qui, dans un moteur de recherche, se renseignent sur les symptômes de la grippe. Et l'on peut établir, en quelques minutes, la cartographie de l'épidémie de grippe. Les données peuvent être comparées dans le temps, dans l'espace. Les séries statistiques peuvent être rapprochées. Le Big Data ne peut pas tout prédire, mais il est un outil nouveau, d'une puissance incomparable, qui change notre manière d'agir. » *Ibid.*, p.17.

ordinateurs munis d'une connexion internet, ou d'un disque dur, étaient en mesure de créer et de collecter des données numériques. Avec l'apparition des téléphones intelligents, et de tous les objets ayant désormais la capacité d'être connectés aux différents réseaux, il est désormais possible d'archiver des informations concernant le monde physique étant donné que ces nouveaux objets connectés nous y accompagnent quotidiennement. Si l'accroissement du volume des données numériques archivées s'accélère, c'est certainement parce que le nombre de SAAD se multiplie grâce à l'émergence d'une offre grandissante d'objets connectés. Conceptualisé comme un réseau reliant l'ensemble des objets interconnectés, et leur permettant d'acheminer l'information recueillie grâce à différents types de capteurs vers des serveurs de stockage²², l'internet des objets consolide le *Big Data* en favorisant la collecte d'un très grand nombre de données représentant mathématiquement le monde physique auparavant difficilement quantifiable. À partir du moment où le stockage massif de données numériques devient lucratif, l'intégration de capteurs automatiques sur les divers objets connectés devient économiquement logique et, dès lors, très largement répandue. En s'immisçant au sein de la vie des gens, les objets connectés ont grandement contribué à la préférence généralisée des grandes compagnies technologiques pour une catégorie à part entière des données numériques, les données personnelles.

1.1.3 Les données personnelles

Comme le souligne Éric Peres, le terme français de *donnée* est trompeur puisqu'il laisse sous-entendre que les informations sont données intentionnellement par les utilisateurs²³. En effet, nous pouvons constater que, dans bien des cas, les usagers des multiples services connectés et réseaux sociaux offerts n'ont pas réellement conscience de l'entière des données personnelles qu'ils laissent derrière eux, ainsi que des renseignements auxquels elles confèrent l'accès. Si les données numériques considérées individuellement peuvent paraître anodines et peu révélatrices, la réalité est tout autre lorsqu'elles sont organisées et structurées en fonction d'identité réelle. Se référant donc à

²² PATEL, Keyur et Sunil PATEL. « Internet of things – IOT : Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges », *International Journal of Engineering Science and Computing*, vol. 6, n°5, May 2016, p. 6122.

²³ PERES, Éric. *Op. cit.*, p. 10.

« toute information concernant une personne physique identifiée ou identifiable »²⁴, les données personnelles sont devenues l'un des « critères principaux de valorisation capitalistique des entreprises de la nouvelle économie. »²⁵ Ce sont elles qui permettent de nombreuses opérations comme la constitution de base de clients, le microciblage publicitaire ainsi que l'ajustement de l'offre à la demande²⁶. Comme le souligne le *Règlement général sur la protection des données*, « les échanges de données à caractère personnel entre acteurs publics et privés, y compris les personnes physiques, les associations et les entreprises, se sont intensifiés dans l'ensemble de l'Union. »²⁷ Manifestement explicable grâce au vaste potentiel de monétisation des données personnelles, la prolifération de ce type de donnée soulève évidemment des questions concernant la frontière de la vie privée. De plus, un nombre important de données permet notamment la prédiction statistique, mais également le profilage lorsque les données analysées sont personnelles. « Désormais, la variété des informations personnelles que nous donnons, souvent involontairement, permet un profilage précis, pour les entreprises privées, de qui nous sommes, de ce que nous pensons, de comment nous vivons. »²⁸ Même si l'individualisation permettant l'identification de la personne est loin d'être l'objectif premier des entreprises détenant nos données personnelles, notre appartenance à des groupes précis et circonscrits représente, quant à elle, une cible ayant une valeur importante lors de contrats publicitaires. Comme nous le verrons au prochain chapitre, le profilage²⁹ généré en fonction des données personnelles suscite des enjeux de justice qui seront centraux pour les considérations de ce mémoire.

²⁴ Journal officiel de l'Union européenne, *Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données)*, 2016, p. 5.

²⁵ PERES, Éric. *Op. cit.*, p. 18.

²⁶ *Ibid.*

²⁷ Journal officiel de l'Union européenne, *Op. cit.*, p. 2.

²⁸ PERES, Éric. *Op. cit.*, p. 56.

²⁹ « [Le profilage désigne] toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique; » Journal officiel de l'Union européenne, *Op. cit.*, p. 33.

1.1.4 L'anonymisation/pseudonymisation des données

Si les données personnelles sont des informations et renseignements permettant l'identification d'un individu, alors ces dernières méritent d'être protégées afin d'assurer la protection de la vie privée comme le stipule le *Règlement général sur la protection des données*³⁰. Le principal moyen utilisé afin d'y parvenir est le processus de pseudonymisation grâce auquel « le traitement des données à caractère personnel [est effectué] de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément. »³¹ Bien que l'anonymisation irréversible des données personnelles soit théoriquement réalisable, plusieurs travaux démontrent que les données personnelles anonymisées peuvent, dans une proportion alarmante des cas, être manipulées en vue d'une réidentification des individus³². La raison de cette réalité est facilement explicable, « data can be either useful or perfectly anonymous but never both. »³³ En effet, si les données personnelles anonymisées confèrent des renseignements sur des personnes physiques identifiables, alors elles ont de la valeur et il devient possible, grâce à certains mécanismes, de les désanonymiser. En contrepartie, les données parfaitement anonymisées sont tout simplement sans valeur puisqu'elles sont dénudées de toutes informations potentiellement utiles. Une étude effectuée aux États-Unis démontre notamment que dans 87% des cas, le code postal, la date de naissance et le sexe sont des informations suffisantes afin d'identifier un individu³⁴. Considérant le nombre astronomique de données personnelles constamment générées par nos actions et comportements connectés, il serait évidemment naïf de croire que les propriétaires de ces dernières, qu'elles soient ou non anonymisées, sont dans l'incapacité véritable de s'infiltrer à l'intérieur des frontières de nos vies privées, « there is only so much anonymity you can promise when recording a person's every move. »³⁵

³⁰ Journal officiel de l'Union européenne, *Op. cit.*, p. 2.

³¹ *Ibid.*, p. 33.

³² OHM, Paul. « Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization », *UCLA Law Review*, 2010, vol. 53, p. 1706-1707.

³³ *Ibid.*, p.1704.

³⁴ SWEENEY, Latanya, *Simple Demographics often Identify People Uniquely*, Carnegie Mellon University, Pittsburgh, 2000, p. 2.

³⁵ KEARNS, Micheal et Aaron ROTH, *The ethical algorithm: the science of socially aware algorithm design*, Oxford University Press, New-York, 2019, p. 2.

1.2 Les algorithmes et le critère d'intelligibilité

Comme nous l'avons présenté, le *Big Data* s'explique par le traitement algorithmique des données numériques et personnelles collectées et archivées qui sont, en elles-mêmes, sans grande valeur. Tout comme les données, les algorithmes précèdent très largement la révolution numérique à laquelle ils sont habituellement associés. Dans sa définition la plus simple, « an algorithm is nothing more than a very precisely specified series of instructions for performing some concrete task »³⁶. Considérant cette définition, l'usage des premiers algorithmes mathématique remonte à l'antiquité et la première systématisation des algorithmes fut effectuée par le mathématicien perse Al-Khwarizmi au début du 9^e siècle³⁷. Architecturé au sein du projet de l'IA, un algorithme désigne désormais « un ensemble d'instructions, une recette construite en étapes programmables, développées dans le but d'organiser et d'agir sur un corpus de données, pour accomplir rapidement un résultat escompté. »³⁸ Avec l'accroissement des percées dans le domaine de l'IA, les processus décisionnels automatisés occupent de plus en plus de terrain à cause des avantages concurrentiels indéniables qu'ils génèrent. L'usage des algorithmes est devenu la norme et le développement de cette technologie est vaste et extrêmement dynamique. Avant d'aborder la question des implications sur la vie des gens, nous nous attarderons sur la classification de différents types d'algorithmes afin d'en clarifier le fonctionnement.

En lui-même, le fonctionnement des algorithmes est fondamental puisqu'il permet une compréhension des mécanismes expliquant concrètement les impacts qu'ils peuvent avoir sur les gens. C'est en décortiquant leurs mécanismes que nous pourrions éventuellement qualifier les effets et les implications de leur action. Pour ce faire, nous diviserons les algorithmes en deux catégories, les algorithmes déterministes et les algorithmes apprenants. Nous porterons une attention toute particulière au concept de l'intelligibilité dont l'importance est cruciale dans la comparaison des différentes catégories d'algorithmes.

³⁶ *Ibid.*, p. 4.

³⁷ Allard, *Muhammad Ibn Musa Al-Khwarizmi, Le calcul Indien (algorismus)*. Versions latines du XIII^e siècle, Librairie scientifique et technique Albert Blanchard, Paris, 1992.

³⁸ Déclaration de Montréal IA responsable, *Op. cit.*, p.286.

1.2.1 Les algorithmes déterministes

La nature des algorithmes déterministes se rapproche énormément des algorithmes mathématiques utilisés avant l'avènement de l'informatique. Leur fonctionnement est en effet déterminé par une démarche procédurale. Tangiblement, lorsque des données sont soumises à un algorithme déterministe, ce dernier effectue une suite d'actions prédéterminées et finies associant systématiquement les sorties (output) aux entrées (input) que sont les données. L'algorithme détient donc un pouvoir décisionnel puisqu'il prendra finalement la décision qui découlera mathématiquement des données lui ayant été fournies. Une multitude d'algorithmes déterministes se déploie au sein de notre réalité quotidienne comme c'est par exemple le cas avec les calculatrices ou encore les algorithmes utilisés pour le triage à l'urgence des hôpitaux que nous expliquerons. Grâce à une pondération numérique de différents critères, un score est attribué aux patients et ce dernier détermine l'ordre de priorité de ces derniers. Les algorithmes déterministes se réduisent donc à des suites d'opérations mathématiques dont l'ordre est programmé en amont et dont l'application à des données permet l'extraction d'une sortie, ou d'un score en l'occurrence. « [...] Every step of the procedure is explicitly specified by its human designers, and written down in a general-purpose programming language such as Python or C++. »³⁹ Si ces algorithmes sont qualifiés de déterministes, c'est parce que l'application de la procédure algorithmique en détermine inmanquablement la sortie. Comparativement aux algorithmes apprenants, les algorithmes déterministes ont le mérite d'être intelligibles, c'est-à-dire que les décisions qu'ils génèrent sont compréhensibles et procéduralement transparentes. Concrètement, un algorithme est intelligible lorsque son fonctionnement est interprétable et concevable par un être humain ayant les connaissances requises pour comprendre le modèle mathématique qui le détermine.⁴⁰ Les algorithmes déterministes sont utiles afin de résoudre des problèmes relativement simples lorsque les résultats peuvent être déterminés mathématiquement, mais ils sont peu performants lorsque vient le temps d'effectuer des opérations plus complexes comme celles de la reconnaissance faciale ou vocale.

³⁹ KEARNS, Micheal et Aaron ROTH. *Op cit.*, p. 6.

⁴⁰ MÉGNY, Mael et Issam IBNOUHSEIN, *Quelle transparence pour les algorithmes d'apprentissage machine?* Archive ouverte pluridisciplinaire, 2018, p. 8.

1.2.2 Les algorithmes apprenants

Ils remplacent progressivement les algorithmes déterministes, car ils s'appliquent à une plus grande diversité de problèmes tout en permettant l'atteinte d'une plus grande performance. Effectivement, la faculté de reconnaître grâce aux perceptions sensorielles est cruciale pour nombre d'activités humaines, et cette capacité est très difficilement mathématisable et donc automatisable. Afin de mettre les choses en perspective, un enfant de deux ans n'aura aucune difficulté à distinguer l'image d'un chat à celle d'un chien tandis qu'un algorithme déterministe en sera tout simplement incapable puisque aucune procédure mathématique simple n'est suffisante afin de distinguer ce genre d'objet. En l'occurrence et comme le soutient l'extrait suivant, l'apprentissage-machine, rendu possible grâce à une quantité importante de données, est nécessaire pour les tâches de reconnaissance, ainsi que pour plusieurs autres. « Machine learning is sometimes considered a form of “self-programming,” since it’s primarily the data that determines the detailed form of the learned model. This data-driven process is how we get algorithms for more human-life task, such as face recognition, language translation [...] »⁴¹ etc. Aujourd'hui, les algorithmes apprenants constituent davantage la règle que l'exception. L'apprentissage machine se divise généralement selon trois principales orientations stratégiques, l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement.

1.2.3 L'apprentissage machine

Premièrement, comme son nom nous l'indique, l'apprentissage supervisé est possible grâce au travail d'étiquetage effectué par un agent humain. C'est en s'entraînant avec un nombre suffisant de couples de données étiquetées, et donc manuellement associé à la bonne valeur par un être humain (à l'image d'un chat ou d'un chien par exemple), qu'un algorithme en apprentissage supervisé parviendra à développer ses propres règles d'analyse⁴² qui lui permettront de reconnaître l'image d'un chien à celle d'un chat. Deuxièmement, en apprentissage non supervisé, un algorithme apprenant sera en mesure, grâce à un nombre X de données lui étant fournies, de déterminer, grâce au *clustering*, les similitudes et les différences entre celles-ci afin de les classifier de façon homogène.⁴³ La

⁴¹ KEARNS, Micheal et Aaron ROTH. *Op cit.*, p. 6.

⁴² ZIMMER, Mathieu, *Apprentissage par renforcement développemental*, Archive ouverte pluridisciplinaire, 2018, p. 9.

⁴³ *Ibid.*

tâche programmée de ce type d'algorithme est donc d'élaborer des règles de classification permettant la formation de groupes dont le degré d'homogénéité varie en fonction des critères de programmation. Une plateforme comme Facebook par exemple, dont le nombre de profils d'utilisateurs est immense et dont la profondeur de ces profils permet une accumulation importante de données personnelles, utilise très certainement ce type d'algorithme afin de classer les utilisateurs. Troisièmement, « l'apprentissage par renforcement (Sutton et Barto, 1998) repose sur l'utilisation de données indirectement étiquetées par des récompenses. »⁴⁴ Une fois plongé dans un environnement précis, l'algorithme en question devra, en tant qu'agent autonome, agir de façon à optimiser la somme quantitative des récompenses pouvant être positives ou négatives. En enregistrant le résultat de chacune de ses expériences, l'algorithme apprendra à prendre les décisions lui permettant de maximiser, toujours davantage, les récompenses lui étant attribuées. En 2018, il a été démontré que la logique pénalité-récompense pouvait être opérationnalisée dans le monde réel en étant appliquée à un algorithme apprenant la conduite automobile.⁴⁵

À ce jour, les algorithmes apprenants se déploient massivement au sein des multiples objets, programmes et applications qui ont désormais intégré notre quotidien. Des objets connectés aux réseaux sociaux auxquels nous participons, de la reconnaissance faciale aux suggestions personnalisées, cette nouvelle génération d'algorithmes se perfectionne grâce aux différentes formes d'entraînement que le nombre grandissant de données archivées rend possible. À la mesure du critère d'intelligibilité mentionné plus tôt, ces algorithmes sont plus complexes et plus performants que les algorithmes déterministes, mais ils sont également moins intelligibles, et donc plus opaques, car leur fonctionnement n'est pas déterminé par leurs programmeurs, mais plutôt grâce aux données avec lesquelles ils apprennent. « The result is that the complicated, automated decision-making that can arise from machine learning has a character of its own, distinct from that of its designer. The designer may have had a good understanding of the algorithm that was used to find

⁴⁴ Déclaration de Montréal IA responsable, *Op. cit.*, Lexique.

⁴⁵ ZAFFAGNI, Marc, « Cette IA a appris à conduire une voiture autonome en 20 minutes », Futura Sciences, mis à jour le 7 septembre 2018, <https://www.futura-sciences.com/tech/actualites/intelligence-artificielle-cette-ia-appris-conduire-voiture-autonome-20-minutes-71942/>, consulté de 22 juillet 2020.

the decision-making model, but not of the model itself. »⁴⁶ Il y a donc une distinction à faire entre les algorithmes d'optimisation programmés par des agents humains et les modèles que mettent en œuvre les algorithmes apprenants. L'apprentissage est alors programmé et compréhensible, mais ce que les algorithmes apprennent « can be complicated and inscrutable, especially when the input data is itself complex and the space of possible models is immense. »⁴⁷ Les algorithmes apprenants se distinguent donc assurément des algorithmes déterministes par la très faible intelligibilité des modèles qu'ils génèrent.

1.3 L'intelligence artificielle aujourd'hui

1.3.1 L'apprentissage profond

Dorénavant, les travaux en IA convergent vers l'apprentissage profond, une branche de l'apprentissage machine inspirée par le fonctionnement du cerveau. « Learning algorithms for layered neural networks models are inspired by the way that neurons communicate with one another and are modified by experience. Inside the network, the complexity of the world is transformed into a kaleidoscope of internal patterns of activity that are the ingredients of intelligence. »⁴⁸ L'articulation de systèmes neuronaux à plusieurs couches permet effectivement à l'apprentissage profond une meilleure performance que l'apprentissage machine. Ces systèmes neuronaux se déploient actuellement dans la société à une grande vitesse étant donné qu'ils sont versatiles et extrêmement performants. De la reconnaissance vocale au développement des voitures autonomes, l'apprentissage profond est un pôle majeur d'investissement et les grandes compagnies technologiques en multiplient les applications⁴⁹. Concrètement, cet apprentissage utilise des réseaux de neurones artificiels à plusieurs niveaux avec pour objectif d'amplifier significativement le nombre de connexions possibles entre les entrées et les sorties. La profondeur de ces couches permet une très grande transformation des données puisque chaque neurone est interconnecté et parce que les réponses neuronales, tout comme les connexions entre celles-ci, sont définies par des poids et des seuils ayant une incidence sur la sortie. De plus, il faut comprendre que les algorithmes n'agissent pas en tant qu'entités séparées les unes des

⁴⁶ KEARNS, Micheal et Aaron ROTH. *Op cit.*, p. 11.

⁴⁷ *Ibid.*, p.10.

⁴⁸ SEJNOWSKI, Terrence J., *The Deep Learning Revolution: Artificial intelligence meets human intelligence*, The MIT Press, Cambridge, 2018, p. 1.

⁴⁹ *Ibid.*

autres, mais plutôt comme faisant partie intégrante de systèmes au sein desquels des collections d'algorithmes travaillent ensemble en créant ainsi l'illusion d'une action unifiée, c'est notamment le cas des plateformes comme Facebook, Twitter, Netflix ou YouTube.⁵⁰ Sans entrer dans le détail mathématique, l'analogie entre le fonctionnement du cerveau et l'apprentissage profond permet effectivement d'expliquer les gains de performance potentiellement entraînés par un algorithme à plusieurs couches.

Ce qui le distingue de l'apprentissage machine traditionnelle, c'est son architecture lui permettant d'utiliser plusieurs couches dans la résolution de problème. Si nous reprenons l'exemple de la distinction entre l'image d'un chien et celle d'un chat, un algorithme à plusieurs couches serait en mesure d'attribuer chaque couche à différentes caractéristiques (par exemple aux contours, aux combinaisons de contours, aux objets modélisés, etc.) de l'image à traiter. Grâce à l'apprentissage profond et au nombre de couches qu'il est possible d'y intégrer, la distinction entre un chat et un chien devient évidemment possible, mais la reconnaissance d'une image aléatoire l'est soudainement également. Dans les faits, les algorithmes en apprentissage profond peuvent s'attaquer à des problèmes plus complexes que les algorithmes apprenants puisqu'ils sont capables de mobiliser une interaction entre un plus grand nombre de critères. Au cours des dernières années, une multitude de recherches se sont intéressées aux applications problématiques de cette nouvelle technologie. Dans ce registre, nous pouvons penser aux algorithmes d'apprentissage profond ayant été développés dans les domaines de la justice digitale, de la police prédictive ou encore de la conduite automatisée. Les décisions adoptées par ces algorithmes peuvent avoir des répercussions importantes sur la vie des gens comme le démontre le célèbre cas de figure du système COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) grâce auquel la firme ProPublica a découvert « that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk.»⁵¹ Un système neuronal

⁵⁰ BUCHER, Tania, « Neither black nor Box », dans BUCHER, Tania (dir.), *If...Then: Algorithmic Power and Politics*, Oxford University Press, New-York, 2018, p. 11.

⁵¹ LARSON, Jeff, MATTU, Surya, KIRCHNER, Lauren et Julian ANGWIN, « How We Analyzed the COMPAS Recidivism Algorithm », *ProPublica*, mis à jour le 23 mai 2016,

comme ce dernier article 137 différents critères (excluant celui de l'ethnicité), dont les poids et les seuils sont variables, afin de déterminer le risque de récidive des individus sur une échelle de 1 à 10.⁵² Lors de cette analyse, il fut alors déterminé qu'en dépit de l'exclusion du critère de l'ethnicité, cet algorithme d'apprentissage profond discriminait les hommes noirs en leur accordant un plus haut taux de récidive. Il était néanmoins impossible pour les programmeurs et opérateurs de cet algorithme d'expliquer la discrimination de ces résultats en raison de l'inintelligibilité de ce dernier.

1.3.2 L'analogie des boîtes noires

Le concept de la boîte noire fait littéralement référence à un objet dont le fonctionnement intérieur n'est pas observable à cause de la noirceur y entravant la vision.⁵³ Au cours des dernières années, l'analogie avec la boîte noire a beaucoup été utilisée à cause de son application aux algorithmes. Dans *The Black Box Society, The Secret Algorithms That Control Money and Information* parut en 2015 et rédigé par Frank Pasquale, il est spécifiquement question de l'opacité des algorithmes qu'il qualifie de boîtes noires inintelligibles. Selon ce dernier, le premier niveau d'explication de cette analogie provient du manque de transparence résultant du caractère confidentiel⁵⁴ des développements effectués en apprentissage machine et profond par les grandes compagnies technologiques. À titre d'exemple, Google défend vigoureusement le complexe d'algorithmes de son moteur de recherche⁵⁵ et pour cette raison, il devient impossible d'avoir accès à l'intérieur de cette boîte algorithmique. Pasquale affirme cependant que même si des conditions de parfaites transparences étaient imposées afin que nous puissions accéder à l'architecture du complexe d'algorithmes de Google, la trop grande complexité de ce dernier assurerait néanmoins son opacité, et donc son statut de boîte noire.⁵⁶ Si la transparence constitue effectivement une condition nécessaire à la lisibilité des algorithmes inintelligibles, elle ne

<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>, consulté le 5 septembre 2020.

⁵² RAHMAN, Farhan, « COMPAS Case Study: Fairness of a Machine Learning Model », *Towards Data Science*, mis à jour le 7 septembre 2017, <https://towardsdatascience.com/compas-case-study-fairness-of-a-machine-learning-model-f0f804108751>, consulté le 5 septembre 2020.

⁵³ BUCHER, Tania. *Op. cit.*, p. 5.

⁵⁴ PASQUALE, Frank, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge, 2015, p. 6.

⁵⁵ *Ibid.*, p.13.

⁵⁶ *Ibid.*, p.8.

constitue pas forcément un moyen pour y arriver. Comme nous l'avons déjà démontré, la grande performance des algorithmes apprenants (incluant les algorithmes en apprentissage profond) s'explique grâce à leur très grande puissance de traitement ainsi qu'au nombre immense de données à leur fondement. C'est pour cette raison que même si nous avons accès au fonctionnement interne de l'algorithme COMPAS par exemple, il est impossible pour un humain d'expliquer objectivement les décisions émises. Il est cependant possible, grâce à une approche heuristique, d'arriver à des raisonnements expliquant possiblement les décisions algorithmiques complexes.⁵⁷ En effet, il serait potentiellement légitime de défendre que le haut taux de récidive attribué aux populations noires par l'algorithme COMPAS soit explicable par des facteurs systémiques dont la provenance sociohistorique est vérifiable.

L'analogie des boîtes noires et des algorithmes complexes ne fait donc pas l'unanimité puisque l'acceptation de cette comparaison anéantirait toutes volontés de compréhension du fonctionnement interne des processus décisionnels automatisés apprenants. Comme le soutient Bucher, ce n'est pas parce l'homme est actuellement dans l'incapacité intellectuelle de comprendre les processus algorithmiques qu'il est théoriquement impossible qu'il le devienne grâce à la mobilisation des ressources nécessaires.⁵⁸ À cet effet, la cartographie des logiques opérationnelles des algorithmes en termes technographiques offre certaines pistes de solutions.⁵⁹ De plus, Bucher démontre que l'analogie des boîtes noires est problématique puisqu'elle ne tient pas compte du caractère dynamique des complexes algorithmiques. Il existe, par exemple, plusieurs permutations (ou versions) de Netflix et il serait, pour cette raison, plus juste de concevoir les algorithmes non pas comme des *Black Box* mais plutôt comme *eventful*.⁶⁰ De cette manière, l'algorithme est défini comme un processus en devenir dont les décisions sont continuellement influencées par les données intégrées.⁶¹ « Algorithms are not given, they are not either mathematical expressions or expressions of human intent but emerge as

⁵⁷ BUCHER, Tania. *Op. cit.*, p. 24.

⁵⁸ *Ibid.*, p.5.

⁵⁹ *Ibid.*, p. 28.

⁶⁰ *Ibid.*, p.12.

⁶¹ *Ibid.*, p.13.

situated ongoing accomplishment. »⁶² Si nous estimons que l’analogie de la boîte noire stérilise vainement les efforts actuellement mobilisés en vue d’un déploiement responsabilisé de l’IA, nous croyons néanmoins que l’inintelligibilité des algorithmes complexes, conceptualisée en tant qu’accomplissement en devenir, est une réalité avec laquelle nous sommes actuellement dans l’obligation de composer.

1.3.3 La compréhension globale des effets

Si nous revenons à l’exemple de l’algorithme COMPAS dont l’inintelligibilité résulte de la complexité et de l’immensité des connexions possibles entre les 137 critères qu’il articule, il est clair qu’il est impossible pour un juge d’expliquer objectivement l’attribution automatisée d’un taux de récidive. Dans un cas de figure comme celui-ci, la conséquence que subit l’individu, en l’occurrence l’allongement ou le raccourcissement de la peine d’emprisonnement, est directe et humainement injustifiable. L’incompréhension d’une décision algorithmique soulève la question de la légitimité des répercussions que provoquent ces nouveaux agents décisionnels. À partir du moment où une décision algorithmique touche négativement certaines personnes à cause de caractéristiques les représentant, nous pouvons déduire qu’il y a discrimination et que les implications de l’algorithme COMPAS créent des injustices. En contrepartie, si l’algorithme en question était intelligible, alors il aurait été possible de déterminer, grâce à une analyse de la programmation, la raison exacte de cette discrimination systémique. Néanmoins, la réalité des algorithmes apprenants et des algorithmes en apprentissage profonds est tout autre puisque leur fonctionnement, ainsi que les résultats qu’ils génèrent, est ultimement influencé par une myriade de facteurs pouvant dépasser l’objet de la programmation. Dans le troisième chapitre, la question des biais algorithmiques sera abordée en profondeur, mais pour l’instant, il est seulement nécessaire d’établir que les processus décisionnels automatisés complexes ont des effets concrets sur nos vies.

Il existe une grande variété d’algorithmes complexes et inintelligibles, et il serait mal avisé d’affirmer que l’entière existence d’entre eux engendre des impacts négatifs pour l’homme. Les avancées technologiques facilitent plusieurs facettes de nos existences, et seulement le perfectionnement de la reconnaissance vocale ou des systèmes de navigation en sont des

⁶² *Ibid.*, p.21.

exemples flagrants. En contrepartie, tout en acceptant les gains potentiels émanant du projet de l'IA, nous souhaitons faire la lumière sur les risques encourus par le déploiement des algorithmes inintelligibles. Le nombre de travaux et d'articles scientifiques cataloguant les multiples cas où des décisions algorithmiques causent des préjudices à des personnes ou à des groupes spécifiques de personnes, ne cesse d'augmenter. « [...] controlled online experiments have demonstrated racial, gender, political, and other types of bias in Google search results, Facebook advertising, and other Internet services. »⁶³ Comme le souligne Safiya Umoja Noble, « we are only beginning to understand the long-term consequences of these decision-making tools in both masking and deepening social inequality. »⁶⁴ Sur ce point, la causalité entre les décisions algorithmiques et leurs effets matérialisés socialement est très loin d'être explicite et intégralement théorisée. Face au nombre incommensurable de processus décisionnels automatisés incessamment à l'œuvre, nous croyons qu'il est nécessaire d'adopter une résolution d'analyse plus large afin d'avoir une vue d'ensemble des effets des algorithmes sur la société. Si les décisions algorithmiques ont quotidiennement des implications sur la vie des hommes, alors le monde de demain est partiellement et activement construit par ces accomplissements en devenir que sont les algorithmes inintelligibles. C'est précisément de cette idée que provient notre hypothèse de l'accroissement des injustices structurelles provoqué par l'utilisation grandissante des processus décisionnels automatisés. Si la justice sociale est altérée par les algorithmes, alors il est nécessaire d'expliquer ce phénomène afin d'être en mesure d'en diminuer les effets. Ce mémoire se positionne formellement dans une volonté de clarification à l'égard de la compréhension globale des effets potentiels des décisions algorithmiques sur nos vies et ultimement, sur la justice sociale.

⁶³KEARNS, Micheal et Aaron ROTH. *Op cit.*, p. 14.

⁶⁴NOBLE, Safiya Umoja, *Algorithms of Oppression: How Search Engines Reinforce Racism*, New-York University Press, New-York, 2018, p. 14.

2 Performativité algorithmique et construction identitaire

Afin de soutenir notre thèse initiale, celle de l'accroissement des injustices structurelles provoquées par la multiplication des outils algorithmiques dans la société, nous prouverons que le traitement de masse des données a une incidence sur la construction identitaire des êtres humains. Pour ce faire, nous introduirons le concept de performativité algorithmique. Ce dernier se présente comme un phénomène qui explique l'effet de la force émergeant de l'accumulation et de la répétition des décisions algorithmiques sur la construction identitaire. Après avoir présenté sommairement le fonctionnement et les implications du déploiement de l'IA au premier chapitre, nous aborderons désormais l'objet de notre thèse.

À cette fin, nous étudierons d'abord le concept de performativité qui provient de l'approche linguistique élaborée par J.L. Austin dans les années soixante, pour ensuite nous pencher sur une réinterprétation de cette notion, théorisée par Judith Butler, afin d'expliquer la construction performative du genre. Grâce à un cadre théorique intersectionnel, nous soutiendrons que l'identité sociale est une construction performative s'articulant à l'intersection des groupes sociaux auxquels appartiennent les individus. Nous pourrons par la suite démontrer que les décisions algorithmiques participent activement à la constitution de cette dernière. En ce sens, nous estimons qu'il est possible, notamment grâce à la notion d'émergence, d'expliquer le caractère performatif des complexes algorithmiques grâce au nombre important de décisions qu'ils génèrent. Il sera finalement possible d'approfondir notre compréhension des effets de la performativité algorithmique sur la construction identitaire en établissant des liens entre, d'une part, l'identité sociale à l'intersection des groupes sociaux la constituant, et d'autre part, l'identité numérique à l'intersection des catégorisations algorithmiques (*clusters*). Nous démontrerons ainsi que la socialisation, en tant que processus participant à la construction identitaire, explique la continuité entre ces deux identités traditionnellement hermétiques.

2.1 La performativité

2.1.1 L'origine du concept

La performativité est un concept qui a grandement voyagé depuis son extraction du cadre linguistique l'ayant vu naître. Effectivement, après qu'Austin ait formulé ses premières intuitions et intentions face à la notion de performativité, il est intéressant de constater que le concept de performativité a circulé entre plusieurs auteurs, objets de

recherches et espaces disciplinaires. C'est dans le livre *How to do Things with Words*, un livre regroupant les douze conférences d'Austin présentées entre les années 1955 et 1960, que l'intention de ce dernier s'est illustrée, c'est-à-dire « [...]définir le caractère performatif de certains énoncés, [et donc de] spécifier les différentes occasions où l'énonciation n'est pas en train de « constater » une action ou une situation, mais constitue elle-même une action à part entière. »⁶⁵ Son projet visait à faire la distinction entre les énoncés constatifs, associés aux affirmations classiques auxquelles une valeur de vérité doit être attribuable, et les énoncés performatifs, qui permettent d'accomplir des actions concrètes se distinguant de leur énonciation.⁶⁶ Afin de clarifier cette séparation, la démarche retenue fut alors de considérer l'entièreté des énonciations destinées à communiquer et de les examiner comme des actes de langage produits par les interlocuteurs.⁶⁷

L'entreprise d'Austin est donc motivée par la distinction entre deux types d'énoncés linguistiques, les constatifs et les performatifs. Premièrement, les énoncés constatifs sont soumis au critère de vérité puisqu'ils sont obligatoirement vrais ou faux. Néanmoins, « toutes les affirmations, vraies ou fausses, ne sont pas pour autant des descriptions; voilà pourquoi [Austin] préfère employer le mot constatif. »⁶⁸ Lorsqu'Austin élabore les premiers critères de la différenciation entre le constatif et le performatif, il est d'abord⁶⁹ clair pour ce dernier que les énoncés constatifs n'accomplissent rien, outre le fait d'être vrais ou faux. La particularité des énoncés performatifs s'établit grâce à cette faculté qu'on certains actes de langage d'accomplir quelque chose par opposition au fait de simplement affirmer quelque chose. Lorsqu'il est question de ce type d'énoncé, « l'énonciation de la phrase est l'exécution d'une action (ou d'une partie de cette exécution) qu'on ne saurait, répétons-le, décrire tout bonnement comme étant l'acte de dire quelque chose. »⁷⁰ En d'autres termes, l'énonciation performative permet d'accomplir quelque

⁶⁵ DENIS, Jérôme. « Préface : Les nouveaux visages de la performativité », *Étude de communication*, OpenEdition Journals, numéro 29, 2006, p. 2.

⁶⁶ LANES, Gilles, Préface dans *Quand dire, c'est faire*, (traduit de l'anglais par Gilles Lanes), Paris, Édition du Seuil, 1970 [1962], p. 24.

⁶⁷ *Ibid.*, p.20.

⁶⁸ *Ibid.*, p.39.

⁶⁹ Austin reconnaîtra, lors de ses analyses ultérieures, un certain *faire* inhérents aux énoncés constatifs.

⁷⁰ *Ibid.*, p.40.

chose grâce à la parole. Même s'il est vrai que l'énonciation d'une parole est effectivement un acte accomplissant l'existence de cette dernière, l'énonciation performative vise néanmoins un accomplissement dépassant cette finalité qui est, en définitive, partagée par tous les actes de discours. Le propre des énonciations performatives réside dans le fait qu'elles accomplissent ce qu'elles sont destinées à accomplir lors de leur énonciation. Malgré le fait que la performativité se limitait initialement à la philosophie du langage, et plus précisément aux actes de discours, il sera intéressant d'explorer le potentiel de ce concept en admettant que les actes de langage puissent prendre d'autres formes pouvant dépasser les limites sémantiques qu'Austin s'était imposées.

2.1.2 La performativité réinterprétée

Comme nous l'avons précédemment indiqué, le concept de performativité a été utilisé à maintes reprises par différents auteurs depuis la première formulation d'Austin. Jérôme Denis attribue le succès de la notion de performativité à un double tournant épistémologique se divisant, d'une part, en « *un tournant linguistique* qui doit beaucoup aux relectures du « second » Wittgenstein (1961) et qui a vu naître un intérêt majeur pour les pratiques langagières et plus généralement la « communication » dans les sphères qui en était étrangères (par exemple l'étude des pratiques scientifiques, celles des activités juridiques, ou encore des sciences de l'organisation [et, d'une autre part, en] *un tournant pratique* (Schatzki et al., 2001) qui a suscité de nombreuses recherches attachées à décrire et à comprendre l'action et les activités qui composent leur objet de recherche. »⁷¹ Dans *The practice turn in contemporary theory*, Theodore R. Schatzki explique ce tournant vers la théorie pratique en affirmant que le « concept of *practice frequently points to repetitive and collective types of conduct* »⁷² afin d'expliquer le monde social et culturel. Pour sa part, Judith Butler s'inscrit très certainement dans cette approche puisqu'elle soutient que certains actes et agissements acquièrent une force performative à cause de leur répétition. Si seulement les actes de discours pouvaient être performatifs selon Austin, la réinterprétation du concept de performativité de Butler stipule cependant que les actes du corps, en terme interactionnel, sont tout autant performatifs. Comme nous le verrons, les

⁷¹ DENIS, Jérôme. « Préface : Les nouveaux visages de la performativité », *Étude de communication*, OpenEdition Journals, numéro 29, 2006, p. 1.

⁷² SCHATZKI, Theodore R., Karin Knorr CETINA et Eike Von SAVIGNY, *Practice turn in contemporary theory*, London, 2001, p. 79.

actes ne sont pas performatifs dans leur accomplissement, c'est plutôt la répétition de ces derniers qui confère une force performative à certaines accumulations d'actes.

2.1.2.1 *La performativité du genre*

L'inspiration première de ce mémoire provient des travaux de Judith Butler, dans le champ des *cultural studies*, qui intègrent une réflexion basée sur la notion de performativité. Dans *Gender Trouble : Feminism and the subversion of Identity* (1990), Butler développe la célèbre proposition de Simone de Beauvoir soutenant que « one is not born, but, rather, becomes a woman. »⁷³ Ce sera grâce à une réinterprétation du concept de performativité que Butler parviendra à définir le genre en soutenant qu'il s'agit effectivement d'une construction sociale comme le suggère initialement Beauvoir. Coordonnée autour de la construction culturelle du genre, la performativité soutient que le genre est un acte⁷⁴, ce qui signifie que l'identité genrée n'existe pas en tant que réalité intérieure objective et qualifiable, elle « consiste davantage en une identité tissée avec le temps par des fils ténus, posée dans un espace extérieur par une répétition stylisée d'actes. »⁷⁵ La construction performative du genre est donc le produit de la performance sociale des individus s'exprimant dans l'accumulation d'« actes, de gestes et de désirs exprimés et réalisés »⁷⁶. L'hétérosexualité reproductive résulte donc de la naturalisation de l'association couplant les performances sociales aux identités genrées existantes, le féminin et le masculin.⁷⁷ Si le genre n'est que performatif, c'est-à-dire qu'il existe uniquement grâce aux performances sociales des individus y adhérant, alors le genre est une construction culturelle ne détenant aucune justification objective, mis à part la pérennité de fictions culturelles. La performance d'une identité genrée est motivée par la permanence d'un soi genré rendu possible grâce à des « actes répétés visant à s'approcher de l'idéal du fondement substantiel pour l'identité, mais qui, à l'occasion de discontinuités, révèlent l'absence, temporelle et contingente, d'un tel fondement. »⁷⁸ Si le genre n'a aucun fondement substantiel à cause de sa nature performative, alors ses fondations sont d'ordres interactionnels, culturels et idéologiques. À chaque fois qu'un homme ou qu'une femme

⁷³ DE BEAUVOIR, Simone, *Le deuxième sexe, Les faits et mythes*, Paris, Gallimard, 1949, p. 301.

⁷⁴ BUTLER, Judith. *Op. cit.*, p. 264.

⁷⁵ *Ibid.*, p. 265.

⁷⁶ *Ibid.*, p. 259.

⁷⁷ *Ibid.*

⁷⁸ *Ibid.*, p. 265.

agit conformément aux représentations hétéronormatives communément acceptées, que ce soit avec une posture, une façon de s'exprimer ou en adoptant une conception du monde, ce ou cette dernier(e) participe à la construction performative du genre. Puisque le genre n'a aucune réalité ancrée dans la genèse sexuée de l'homme, il devient donc juste d'affirmer que le genre est dynamique et malléable, qu'il se transforme, et qu'il est modifié par l'agissement de ses représentants. Inversement, il faut également comprendre qu'il existe un jeu interactionnel entre le genre, en tant que construction performative, et ses représentants, en tant qu'incarnations de cet objet mouvant. Effectivement, le genre performatif devient un point de référence influençant performativement les agissements des êtres humains. Ce va-et-vient incessant est une fiction régulatrice permettant l'illusion d'une identité de genre originelle.⁷⁹

Concrètement, si le genre est le produit des performances sociales incarnées par les représentants de l'hétéronormativité, alors la construction identitaire est partiellement performative puisqu'elle s'effectue grâce au revêtement des attributs et des actes de genre découlant de ces actes de performances répétées. Bien que Butler se soit limitée à la construction performative du genre, nous tenterons d'étendre le potentiel du concept de la performativité à la construction identitaire.

2.1.2.2 *La métaphore théâtrale*

« This wide and universal theatre
Presents more woeful pageants than the scene
Wherein we play in. »⁸⁰

Pouvant être traduite de la manière suivante : « *Le monde entier est un théâtre, et tous, hommes et femmes, n'en sont que les acteurs. Et notre vie durant nous jouons plusieurs rôles.* », cette citation de Shakespeare résume l'essentiel de la théorisation de la métaphore théâtrale d'Erving Goffman. Représentant de l'école de Chicago, ce dernier s'inscrit dans le courant de l'interactionnisme symbolique, approche sociologique se fondant sur l'idée que la société, ou plus précisément la réalité sociale, est le produit des interactions entre les individus.⁸¹ Dans *The Presentation of Self in Everyday Life* parut en

⁷⁹ *Ibid.*, p. 266.

⁸⁰ SHAKESPEARE, William, *As you like it*, Londres, 1599, p. 45.

⁸¹ BERGER, Peter L. et Thomas LUCKMANN, *La construction sociale de la réalité*, (traduit de l'anglais Pierre Taminioux), Armand Colin, 2014, [1966], p. 213.

1956, Goffman ira plus loin en affirmant que l'identité découle également de l'interactionnisme symbolique. Afin de soutenir cette proposition, il développe l'idée que la *présentation de soi* symbolise le rôle, ou la performance, qu'incarnent les individus dans le théâtre quotidien de la vie.⁸² En stipulant que toutes actions sociales s'apparentent à une performance théâtrale, les acteurs, ou individus, chercheront à obtenir des informations afin d'adapter adéquatement leurs *présentations de soi* aux différentes situations interactionnelles composant nos existences. En ce sens, nos performances devraient avoir pour objectif, dans la majorité des cas⁸³, de se conformer aux attentes sociales encourues. Au cours d'une vie, un individu doit donc être en mesure d'incarner plusieurs performances s'il ne veut pas perdre la face, ou être démasqué⁸⁴, c'est-à-dire s'il ne veut pas que son jeu soit dévoilé, et qu'une incohérence identitaire soit révélée. Bien entendu, ces performances se chevauchent dans un jeu interactionnel et l'analogie du rôle du professeur et de l'étudiant est souvent employée dans la littérature afin d'illustrer cette réalité. Sur le plan social, différents attributs/agissements sont attendus d'un élève et d'un professeur, et lorsqu'un individu incarne l'un de ces deux rôles, sa performance doit se conformer à ces exigences. Néanmoins, tous les professeurs ont déjà été étudiants, certains sont simultanément les deux à la fois, et une multitude d'autres rôles s'immiscent inévitablement dans la *présentation du soi*, que ce soit ceux associés au genre, au sexe ou encore à l'ethnicité.

Erving Goffman fait la distinction entre deux types d'acteurs, le sincère et le cynique. Le sincère est celui dont la performance est incarnée véritablement et inconsciemment, tandis que le cynique est celui ayant pertinemment conscience du travestissement volontaire de ses performances.⁸⁵ Cette distinction présente plusieurs difficultés puisqu'elle nous mène à la question de l'arrière-scène, symbolisant le retraitement d'un individu de toutes interactions sociales, et pouvant faire référence à une identité substantielle et objective. Un acteur cynique a-t-il continuellement accès à son identité véritable puisqu'il est conscient de la performance qu'il joue? L'identité d'un acteur sincère existe-t-elle uniquement dans l'articulation des performances jouées

⁸² GOFFMAN, Erving, *The presentation of the self in everyday life*, Écosse, 1956, p. 33.

⁸³ Goffman souligne qu'il existe des cas de rebellions et de révolutions expliquant que certains individus préfèrent aller volontairement à l'encontre des conventions sociales existantes.

⁸⁴ GOFFMAN, Erving, *Op. cit.*, p. 40.

⁸⁵ *Ibid.*, p. 10.

inconsciemment? Pour répondre à ces questionnements, il est intéressant de constater que dans son livre de 1990, Judith Butler emprunte cette analogie : « [...] une identité construite, un acte performatif que le grand public, y compris les acteurs et actrices elles/eux-mêmes, viennent à croire et à reprendre [*perform*] sur le mode de la croyance. »⁸⁶ Butler poussera plus loin le raisonnement de Goffman en affirmant que la performance joue un rôle décisif dans la construction du genre étant donné que les interprètes finissent par y croire, c'est-à-dire qu'ils deviennent ultimement ce qu'ils performent. En ce sens, l'existence même de l'arrière-scène, ainsi que la distinction entre les types d'acteurs, perd tout son sens puisque les individus finiront par devenir ce qu'ils incarnent, qu'ils soient seuls dans leur chambre, ou qu'ils aient conscience du rôle qu'ils interprètent. Il est important de rappeler ici que ce phénomène de performativité, tel que réinterprété par Butler, s'applique uniquement à l'identité genrée, mais nous reviendrons assurément sur l'élargissement de cette notion. Avant d'y arriver, nous allons revenir sur la notion de performance puisqu'elle sera essentielle dans la théorisation de la performativité algorithmique.

2.1.2.3 *La performance*

Concrètement, et selon les besoins théoriques de ce mémoire, la performance représente le médium entre les algorithmes et l'identité, c'est-à-dire ce qui explique la performativité algorithmique. Dans une transaction bidirectionnelle, les performances générées par la répétition des décisions algorithmiques construisent performativement l'identité des individus, et l'accumulation des performances identitaires construit performativement les identités de groupe, ou les groupes sociaux. Si la compréhension de ce va-et-vient est primordiale pour saisir l'essence du phénomène soulevé ici, les différentes orientations de la performance seront scindées à des fins explicatives. Au prochain point, il sera question des performances engendrées par l'action des algorithmes sur les individus, ces dernières devenant constitutives de l'identité. Nous aborderons également l'influence des performances collectives sur l'identité de groupe ainsi que sur la catégorisation algorithmique.

⁸⁶ BUTLER, Judith. *Op. cit.*, p. 265.

En ce qui a trait au développement de la notion de performance, nous nous inspirerons du champ d'études interdisciplinaire des *Performance Studies*. Cette approche soutient que la performance doit être utilisée comme une lentille et un outil pour étudier le monde. Si dans son sens le plus large, la performance représente toutes actions effectuées dans le monde social,⁸⁷ l'Université Brown définit l'objet d'étude des *Performance Studies* comme « the multiple modes in which live performance articulates culture, negotiates difference, constructs identity, and transmits collective historical traditions and memories. »⁸⁸ Cette définition succincte de la performance permet toutefois de supposer une interaction performative entre la performance sociale dans le quotidien et la construction identitaire. Plus précisément, les actes et les gestes que les individus accomplissent dans les interactions de la vie de tous les jours constituent des performances telles que nous l'entendons :

« Performances can be either “make-belief ” or “make-believe.” The many performances in everyday life such as professional roles, gender and race roles, and shaping one’s identity are not make-believe actions (as playing a role on stage or in a film most probably is). The performances of everyday life “make belief ” – create the very social realities they enact. In “make-believe” performances, the distinction between what’s real and what’s pretended is kept clear. »⁸⁹

Comme la citation le soulève, il existe une distinction entre les performances incarnées (*make-belief*), et les performances jouées (*make-believe*). Pour les fins de ce mémoire, nous nous limiterons à la première dimension de la performance puisqu'elle participe activement à la construction identitaire comme c'est d'ailleurs le cas pour la performance genrée théorisée par Butler. Cette dernière construit performativement l'identité de genre, et par extension, le genre en lui-même. Tout comme notre définition de la performance le suppose, nous allons dorénavant soutenir que l'identité résulte des performances incarnées par les individus. Si l'identité sociale est plus complexe que l'identité genrée de Butler, c'est grâce au concept de l'intersectionnalité que nous expliquerons le glissement qui nous permettra de théoriser la construction performative de l'identité.

⁸⁷ SCHECHNER, Richard, *Performance studies: An introduction*, New-York, 2002, p.1.

⁸⁸ En ligne au: brown.edu/Departments/Theatre_Speech_Dance/about/performancestudies.html consulté le

⁸⁹ SCHECHNER, Richard, *Op. cit.*, p.43.

2.2 La construction performative de l'identité

2.2.1 Les groupes sociaux et l'identité performative

Au chapitre trois, il sera notamment question de la justification de la conception de justice que nous prioriserons afin de juger l'influence de la performativité algorithmique sur la construction identitaire des individus. Pour des raisons qui seront donc élaborées ultérieurement, nous privilégierons les normes de justices mises de l'avant par Iris Marion Young dans son ouvrage de 1990, *Justice and the politics of difference*. Si nous abordons prématurément ce choix théorique, c'est parce que les notions du groupe social et de l'impérialisme culturel qui sont centrales chez Young seront dès lors mobilisées afin d'expliquer la construction performative de l'identité.

Concrètement, « a social group is a collective of persons differentiated from at least one other group by cultural forms, practices, or way of life. »⁹⁰ Cette citation nous permet effectivement de constater que ce qui différencie fondamentalement les groupes sociaux, c'est ultimement la performance, telle que définie précédemment, des membres constituant ces mêmes groupes. En incarnant une performance, un individu agit en conformité avec les *cultural forms, practices, or way of life* distinguant les groupes sociaux. De plus, Young souligne également le caractère dynamique et malléable de ces groupes sociaux. « Group identity is constructed from a flowing process in which individual identify themselves and others in terms of groups, and thus group identity itself flows and shifts with changes in social process. »⁹¹ Il existe un agencement possible entre le concept de groupe social de Young, et celui de la performativité de Butler étant donné que similairement au genre qui est objectivement et substantiellement inexistant chez Butler, le groupe social de Young est dénué de fondement fixe. En ce sens, ces deux ensembles sont perpétuellement en construction et la notion de performativité devrait selon nous expliquer cette réalité. Young précise d'ailleurs que ce qui permet finalement l'existence des groupes sociaux, ainsi que leurs spécificités, ce sont les interrelations les unissant. « Groups are an expression of social relations; a group exists only in relation to at least one other group. »⁹² Encore une fois, un parallèle est à faire entre la conception de groupes sociaux de Young, et la notion de performativité chez Butler, puisque dans chacun des cas, l'interactionnisme symbolique

⁹⁰ YOUNG, Iris Marion, *Justice dans the politics of difference*, Princeton classics, 1990, p. 43.

⁹¹ *Ibid.*, p. 172.

⁹² *Ibid.*, p. 43.

explique la construction des identités de groupes/identités de genre. En effet, les interactions entre les individus ainsi qu'entre les groupes sont ultimement des performances, et ces dernières participent activement à la construction de l'identité de groupe, et donc de l'identité sociale.

Lorsque Young soutient que les groupes existent dans leurs relations aux autres groupes, il faut comprendre que ces relations reconduisent et expliquent les rapports de force entre ceux-ci. Pour reprendre le genre performatif de Butler, les relations/interactions entre le groupe masculin et le groupe féminin sont historiquement asymétriques dans le sens qu'elles avantagent un groupe au détriment d'un autre. Cette asymétrie performativement construite s'est institutionnalisée et est devenue structurante et normative. L'hétéronormativité suppose effectivement que l'hétérosexualité, ainsi que les rapports de domination du masculin sur le féminin, devient la norme. Pour Young, l'impérialisme culturel englobe cette force normative, et il serait par exemple juste d'affirmer que le genre masculin est une forme d'impérialisme culturel dans le sens de la citation suivante. « Cultural imperialism involves the universalization of a dominant group's experience and culture, and establishment, as the norm. [...] The dominant group reinforces it's position by bringing the other groups under the measure of its dominant norms. »⁹³ Effectivement et même si cet impérialisme culturel est de plus en plus contesté, il est indéniable que les hommes sont privilégiés à cause de normes les avantageant au détriment des femmes. Young soutient que l'identité de groupe est le produit de l'exercice objectivant de l'impérialisme culturel⁹⁴ sur les autres groupes y étant soumis. L'impérialisme culturel agit donc comme une valeur stable à laquelle l'identité des autres groupes est finalement comparée. Cela engendre un sentiment d'affinité⁹⁵ entre les membres d'un même groupe puisque ces derniers subissent à l'unisson les effets de domination et d'oppression de l'impérialisme culturel. Les expériences similaires se convertissent alors en sentiment d'appartenance et de cette affinité s'ensuit l'édification d'un langage et d'une perception caractérisant une situation particulière et distinctive. Dans une majorité de cas, la mobilisation des représentants de ces groupes face aux répercussions

⁹³ *Ibid.*, p. 59.

⁹⁴ *Ibid.*, p. 172.

⁹⁵ *Ibid.*

de l'impérialisme culturel prend la forme de mouvements sociaux prônant la défense d'idéaux. « While their ideas of women's culture, Afro-American culture, and American Indian culture rely on past cultural expressions, to a significant degree these movement have self-consciously constructed the culture that they claim defines the distinctiveness of their groups. »⁹⁶ La réalité culturelle d'un groupe est donc collectivement définie par le passé partagé de ce dernier ainsi que par les idéaux le dirigeant.

En reprenant l'exemple de la construction performative du genre théorisé par Butler, nous observons que les identités masculines et féminines sont construites performativement grâce à une « répétition stylisée d'actes. »⁹⁷ Considérant le genre comme un groupe social dont l'identité de groupe est dynamique, il serait également juste d'affirmer que l'identité collective de tous les groupes sociaux est également dynamique puisqu'elle est construite performativement grâce à l'accumulation des actes de performance coordonnés de leurs membres. Effectivement, les hommes et les femmes, en tant que groupes sociaux dont le sentiment d'appartenance et d'affinité en assure la distinction, participent activement à la construction performative de leur groupe à chaque fois qu'ils et qu'elles agissent en conformité avec leur genre. Dans cette optique, nous soutenons qu'il est juste d'affirmer que la construction identitaire des représentants de tous les groupes sociaux est performative puisque ceux-ci sont des constructions performatives dont la pérennité est soumise à la répétition des actes de performances coordonnés de leurs membres. Il en va de même pour les groupes sociaux représentant l'impérialisme culturel puisque la dominance normative des groupes est également dynamique et malléable, et elle dépend ultimement d'un jeu interactionnel entre l'ensemble des groupes sociaux. L'enjeu des rapports de forces entre les groupes sociaux soulèvera des questions de justice qui seront abordées au troisième chapitre. Il sera notamment question des effets injustes de l'impérialisme culturel sur la construction identitaire des individus appartenant aux groupes sociaux marginalisés.

2.2.2 L'intersectionnalité

À l'évidence, les groupes sociaux ne représentent pas des ensembles monolithiques au sein desquelles leurs membres sont cloisonnés. Un individu peut effectivement

⁹⁶ *Ibid.*

⁹⁷ BUTLER, Judith, *Op. cit.*, p. 265.

s'inscrire dans plusieurs groupes sociaux à la fois puisque ces derniers ne sont pas représentatifs de son intégralité identitaire. Le concept de l'intersectionnalité intervient précisément à la jonction de cette complexité identitaire.

« L'intersectionnalité renvoie à une théorie transdisciplinaire visant à appréhender la complexité des identités et des inégalités sociales par une *approche intégrée*. Elle réfute le cloisonnement et la hiérarchisation des grands axes de la différenciation sociale que sont les catégories de sexe/genre, classe, race, ethnicité, âge, handicap et orientation sexuelle. L'approche intersectionnelle va au-delà d'une simple reconnaissance de la multiplicité des systèmes d'oppression opérant à partir de ces catégories et postule leur interaction dans la production et la reproduction des inégalités sociales. »⁹⁸

À l'origine, le concept de l'intersectionnalité provient du mouvement afroféministe et se limitait spécifiquement à l'intersection du racisme et du sexisme ainsi qu'à l'oppression particulière qui en découle.⁹⁹ L'opérationnalisation de ce concept s'est depuis très largement étendue et les différentes catégorisations sociales qui s'y sont intégrées sont multiples. De plus, l'intersectionnalité transcende désormais la théorisation des inégalités sociales pour s'intéresser à la constitution des identités. Pour l'objet du présent mémoire, nous nous intéresserons à ce pan identitaire de l'intersectionnalité étant donné qu'il permet de rendre compte des individus ainsi que de leur appartenance à de multiples groupes sociaux. En effet, la construction performative de l'identité articulée autour du concept de groupe social nécessite l'intégration de cette dimension intersectionnelle pour être cohérente.

Ayant pénétré l'univers de la psychologie sociale, l'intersectionnalité telle que théorisée par Warner et Shields reconnaît que toute identité sociale, que leurs positions soient privilégiées ou marginalisées, existe à l'intersection des groupes les constituant.¹⁰⁰ « As a construct, intersectionality captures the idea that social identities, identities that stem from group membership (Ashmore et al. 2004), are organizing features of social relationships, and these social identities mutually constitute, reinforce, and naturalize one another (Collins 1990; Crenshaw 1991; Nakano Glenn 1999; Anthias and Yuval-Davis

⁹⁸ BILGE, Sirma, *Théorisation féministe de l'intersectionnalité*, Presses Universitaires de France, 2009, p. 70.

⁹⁹ *Ibid.*, p. 74.

¹⁰⁰ WARNER, Leah R. et Stephanie A. SHIELDS, « The intersections of Sexuality, Gender, and Race: Identity Research at the Crossroads », *Sex Roles*, Vol. 68, New-York, 2013, p. 804.

1983), creating both oppression and opportunity for the individual (Baca Zinn and Thornton Dill 1996). »¹⁰¹ Selon cette approche, même si l'individu vit une expérience cohérente des intersections constituant son identité sociale, il n'en demeure pas moins que ces intersections s'organisent dans un réseau complexe de relations de pouvoir opposant ou unifiant les groupes sociaux en question.¹⁰² À l'évidence, les multiples identités sociales au fondement de la subjectivité d'un individu ne peuvent être considérées comme des entités distinctes puisqu'une telle fragmentation du sujet serait inadéquate pour rendre compte des réalités sociales individuelles. Comme nous le démontrerons à l'instant, le cadre théorique présenté par Warner et Shields nous permet permettant d'élargir la construction performative de l'identité genrée de Butler à celle de l'identité sociale.

Pour récapituler, cette réinterprétation du concept de performativité nous permet d'admettre que l'identité sociale est une construction performative. À l'image de l'identité de genre qui est une fiction inexistante objectivement, l'identité racisée (en tant qu'exemple d'identité associée à un groupe social selon la théorisation de Young) est également une construction sociale dont les fondements sont performativement construits, et qui résulte donc de la répétition d'actes de performance coordonnés par les représentants du groupe en question. Si des conditions d'existence similaire contribuent initialement à regrouper les individus en groupes sociaux, la redéfinition dynamique de ces groupes est performative puisque leur réalité est définie grâce à l'orchestration des actes de performance de leurs membres. Inversement, l'identité sociale d'un individu s'accomplit performativement lorsqu'il incarne une performance sociale, c'est-à-dire une « répétition stylisée d'actes »¹⁰³. Comme le stipule Butler, «de tels actes, gestes et accomplissements, au sens le plus général, sont *performatifs*, par quoi il faut comprendre que l'essence ou l'identité qu'ils sont censés refléter sont des *fabrications*, élaborées, et soutenues par des signes corporels et d'autres moyens discursifs. »¹⁰⁴ De plus, la performativité doit être réfléchié dans un cadre théorique admettant l'intersectionnalité identitaire. Effectivement, si la construction de l'identité sociale d'un individu est performative, elle doit également être intersectionnelle puisqu'elle constitue un amalgame cohérent des différents groupes sociaux auxquels un

¹⁰¹ *Ibid.*, p. 803.

¹⁰² *Ibid.*, p. 804.

¹⁰³ BUTLER, Judith, *Op. cit.*, p.265.

¹⁰⁴ BUTLER, Judith, *Op. cit.*, p. 259.

individu appartient. Concrètement, chaque identité sociale est définie à l'intersection des groupes auxquels appartient l'individu, et ces groupes sociaux sont des constructions qui s'accomplissent grâce au va-et-vient performatif entre les membres et les identités de groupes. Comme nous le démontrerons, cette interprétation théorique de la construction identitaire nous permettra d'établir un rapprochement avec les mécanismes desquels résulte la génération des identités numériques, c'est-à-dire les diverses méthodes de classification algorithmique.

2.2.3 L'identité numérique

Les différents contextes et situations dans lesquels interagit un individu influencent les rôles qu'il arbore ainsi que la présentation qu'il fait de son identité. La socialisation des individus leur permet d'effectuer des choix relativement éclairés quant au dévoilement des multiples facettes de leur identité sociale, mais lorsqu'il est question de l'identité numérique, son contrôle est d'une tout autre nature. À vrai dire, le calcul de la vie privée d'une identité numérique est difficilement concevable et applicable pour nombre d'individus étant donné que la construction de ces identités est implicite, opaque et complexe. Dans les faits, « l'identité numérique est constituée de la somme des traces numériques se rapportant à un individu ou à une collectivité : des traces profilaires¹⁰⁵ correspondant à ce que je dis de moi (qui suis-je); des traces navigationnelles qui renseignent sur les sites que je fréquente et sur lesquels je commente ou j'achète (comment je me comporte); enfin des traces inscriptibles et déclaratives – ce que je publie sur mon blogue par exemple – qui reflètent directement mes idées et mes opinions (ce que je pense). »¹⁰⁶ En réalité, toute utilisation d'objets interconnectés est susceptible de laisser une empreinte identitaire qui sera enregistrée et qui pourra potentiellement être associée à une identité numérique susceptible d'être exploitée de manière profitable. Afin d'être performants dans leur fonction de prédiction, de classification, de catégorisation et de ciblage, les algorithmes doivent avoir accès à des bases de données associées directement, ou indirectement, à des profils d'utilisateurs et donc à des identités numériques. Au fur et à mesure que l'internet des objets interrelie un nombre grandissant de SAAD, la profondeur

¹⁰⁵ Néologisme qualifiant les données personnelles que génèrent les profils d'utilisateurs.

¹⁰⁶ ERTZCHEID, Olivier, *Qu'est-ce que l'identité numérique? Identité numérique et e-réputation*, OpenEdition Press, Marseille, 2013, p. 15.

ainsi que la complexité des profils tend à s'accroître avec les nouvelles pratiques qu'ils rendent possibles.

2.2.3.1 Le « clustering »

Tout comme l'identité sociale des individus est une construction performative amalgamant leurs appartenances intersectionnelles à des groupes sociaux différents, l'identité virtuelle de ces derniers est de la même manière située à l'intersection des classifications algorithmiques dans lesquelles ils s'inscrivent. Fondamentalement, la classification algorithmique consiste en un regroupement d'ensembles de données, en fonction de leurs similitudes et différences, grâce à des mécanismes de *clustering*¹⁰⁷. Lorsqu'appliquée à des profils complexes et détaillés d'utilisateurs comme ceux que génèrent actuellement les réseaux sociaux, la classification algorithmique forme des *clusters*, ou des regroupements de profils d'utilisateur homogènes, qui peuvent ultimement se réduire à des groupes sociaux virtuels. Tout comme les individus dont l'identité sociale est représentée par l'appartenance intersectionnelle à plusieurs groupes sociaux, l'identité virtuelle associée à ces mêmes individus est représentée par l'appartenance intersectionnelle à plusieurs *clusters* générés par la classification algorithmique. Similairement, l'identité virtuelle d'un individu s'inscrit grâce aux traces numériques des activités connectées qu'il exécute dans les multiples *clusters* représentant potentiellement l'entièreté des facettes du prisme de son individualité (âge, sexe, ethnicité, situation économique, position géographique, etc.). Si nous souhaitons être en mesure de théoriser l'influence du phénomène de la performativité algorithmique sur la construction identitaire des individus, nous devons assurément approfondir le lien entre l'identité sociale et l'identité virtuelle. Nous soutenons que ces deux identités parallèles constituent des constructions performatives dont les caractéristiques se déterminent à l'intersection des groupes sociaux/*clusters* auxquels elles appartiennent. En ce sens, nous introduirons le dilemme de la vie privée, *the privacy dilemma*, qui nous permettra d'explorer l'interdépendance l'identité sociale et virtuelle.

¹⁰⁷ ZIMMER, Mathieu, *Apprentissage par renforcement développemental*, Archive ouverte pluridisciplinaire, 2018, p. 9.

2.2.3.2 Le « *privacy dilemma* »

Dans la littérature anglophone traitant des enjeux éthiques de l'IA, la notion de *privacy* est capitale puisqu'elle fait référence au dilemme opposant la *privacy* à la *publicity*¹⁰⁸, ou si nous traduisons, de l'intimité à la visibilité. Théoriquement, ce dernier stipule la chose suivante : « [...] the very nature of identity and social relation works against protecting personal data; this is the privacy dilemma. »¹⁰⁹ Considérant cette opposition, nous croyons que la traduction des termes *intimité* et *visibilité* est sémantiquement insuffisante. Ces notions renvoient à des états et non à des processus, et elles n'expliquent donc pas les mécanismes sociaux au fondement de ce dilemme. Effectivement, l'extrait suivant illustre bien la complexité du concept de *privacy* : « Healthy social interaction depends on effectively interpreting a social situation and knowing how to operate accordingly. This along with an understanding of how information flow, is central to the process of privacy. »¹¹⁰ Lorsque nous interagissons socialement, nos dévoilements identitaires dépendent généralement du rapport contextuel dans lequel s'inscrivent nos interactions sociales.¹¹¹ La notion de *privacy* fait directement référence à ce calcul social permettant d'ajuster le degré de dévoilement de la subjectivité dans un contexte donné, tandis que l'intimité est plutôt caractéristique de l'effet d'une dissimulation identitaire. Dans les faits, un individu orchestre ses interactions sociales en évaluant le degré d'intimité et de visibilité qu'il doit entretenir avec les membres de sa famille ou encore avec ses collègues de travail. Bien entendu, certaines informations identitaires doivent être intimes selon certains contextes et visibles pour d'autres, et la notion de *privacy* explique cette capacité de choix individuel traditionnellement acquise lors de la socialisation. Pour des fins de traduction, l'usage du terme *vie privée* se transposera à celui de *privacy* car la protection de la vie privée fait référence à ce calcul individuel articulant l'intimité et la visibilité.

¹⁰⁸ BOYD, Dana, *Privacy and Publicity in the Context of Big Data*, 2010, p. 10.

¹⁰⁹ LAHLOU, Saadi, « Identity, social status, privacy and face-keeping in digital society », *Social Science Information: Cognitive technologies*, 2008, p. 309.

¹¹⁰ BOYD, Dana, *Op. cit.*, p. 11.

¹¹¹ *Ibid.*

2.2.3.3 Les réseaux sociaux et la dynamique « expressiviste »¹¹²

La fonction des réseaux sociaux dans la complexification des identités numériques est non négligeable puisqu'elle justifie idéologiquement le partage des informations étant auparavant considérées comme privées. « En effet, le paradoxe de ce genre de site est qu'il est nécessaire de les nourrir en donnant des informations sur soi pour augmenter la capacité relationnelle des profils. »¹¹³ En tête d'affiche des réseaux sociaux se retrouve la plateforme Facebook articulant les interrelations de près de 3 milliards d'utilisateurs dans le monde¹¹⁴. Sous cette même bannière se retrouve également Instagram, ainsi qu'une plus récente innovation, Méta, qui pourrait potentiellement révolutionner, et du même fait, exacerber la dynamique expressiviste, ainsi que la collecte de données personnelles y étant associée. En accumulant les informations nécessaires à la complexification des profils, cette entreprise a aggloméré les données personnelles définissant les individus à l'origine de ces identités numériques. Comme nous l'avons mentionné précédemment, un individu socialisé camouflera ou exhibera certaines caractéristiques identitaires dépendamment du contexte social auquel elles sont destinées. Il s'agit de la *privacy* ou du calcul de la vie privée. Considérant les réseaux sociaux, il devient de plus en plus difficile pour les individus d'évaluer justement le contexte réel au sein duquel ils divulguent leurs informations et encore moins d'anticiper les potentiels effets de ces divulgations. De plus, à l'image des relations humaines traditionnelles, les relations virtuelles sur les réseaux sociaux nécessitent un dévoilement minimal de part et d'autre des acteurs concernés. La logique intrinsèque des réseaux sociaux favorise et légitimise donc le dévoilement identitaire des individus en provoquant évidemment un approfondissement des identités numériques. L'architecture même de réseaux sociaux est généralement favorable à ce pullulement des données personnelles puisque l'usage des plateformes sociales est souvent enrichi grâce à l'étalement de préférences et de goûts personnels. S'ajoutant à cette caractéristique fondamentale des réseaux sociaux se greffe une dynamique expressiviste stipulant que les « plateformes relationnelles constituent des espaces de très grande

¹¹² ALLARD, Laurence, et Olivier BLONDEAU, « L'activisme contemporain : Défection, expressivisme, expérimentation », *Rue Descartes*, numéro 55, 2007, p. 111.

¹¹³ *Ibid.*

¹¹⁴ En ligne au : <https://www.journaldunet.com/ebusiness/le-net/1125265-nombre-d-utilisateurs-de-facebook-dans-le-monde/#:~:text=Fin%202020%2C%20Facebook%20comptait%20,rapport%20au%20premier%20trimestre%202020> , consulté le 22 juin 2021.

exposition pour les personnes qui s’y engagent. »¹¹⁵ En effet, les réseaux sociaux sont des espaces favorisant l’ « expression de la singularité irréductible de la subjectivité »¹¹⁶ et plusieurs nouvelles modalités de subjectivation sont liées aux possibilités qu’offrent ces plateformes.¹¹⁷ Les réseaux sociaux encouragent les prises de positions idéologiques et nombre d’utilisateurs apaisent leur volonté d’affirmation grâce à ces nouveaux médiums au sein desquels ils se sentent écoutés et compris. « [...] L’identité numérique est une coproduction où se rencontrent les stratégies des plateformes et les tactiques des utilisateurs. » En des termes plus simples, les réseaux sociaux sont des plateformes idéales pour récolter et schématiser des données personnelles afin de les ériger en identité numérique, identités finalement construites grâce aux activités numériques, ou performances, des individus. Dans les faits, la très grande prolifération des réseaux sociaux est justifiée économiquement à cause de la qualité, et donc de la valeur, des données personnelles qu’ils parviennent à récolter.

2.2.3.4 *Les notions de granularité, porosité et percolation*¹¹⁸

Les identités numériques transcendent désormais ces réseaux et afin d’explicitier ce phénomène, nous aborderons trois principes expliquant le cheminement des données personnelles soit la granularité, la porosité et la percolation¹¹⁹. La granularité suppose que « toute activité, même minime, même supposée confidentielle, constitue aujourd’hui pour les moteurs de recherche un document au sens propre. »¹²⁰ Effectivement, l’utilisation du Web et de ses services implique nécessairement que chacun laisse derrière des empreintes numériques. Bien que ces traces soient distinctes et sans valeur à l’image des granules, l’accumulation de ces fragments d’identité numérique est cruciale pour leur élaboration. Un like sur Facebook, un achat sur Amazon, une recherche sur Google ou encore un commentaire sur YouTube sont tous des représentants granulaires d’activité ayant eu lieu sur le Web. Pour plusieurs, le caractère disparate de ces activités devrait être suffisant pour troubler l’édification d’une identité numérique identifiable, mais la réalité dément cette

¹¹⁵ CARDON, Dominique, « Le design de la visibilité, Un essai de cartographie du web 2.0 », *Réseaux*, numéro 152, 2008, p. 97.

¹¹⁶ ALLARD, Laurence et Olivier BLONDEAU, *Op. cit.*, p. 53.

¹¹⁷ *Ibid.*

¹¹⁸ ERTZSCHEID, Olivier, *Qu’est-ce que l’identité numérique, Identité numérique et e-réputation*, OpenEdition Press, Marseille, 2013. p. 47.

¹¹⁹ ERTZSCHEID, Olivier, *Op. cit.*, p. 47.

¹²⁰ *Ibid.*

présomption. À l'inverse de la construction de l'identité sociale qui résulte d'un calcul circonstanciel expliqué précédemment, la construction de l'identité numérique regroupe la totalité des traces numériques granulaires qu'un individu génère grâce à ses activités connectées. Ultiment, les différentes performances enregistrées convergeront en une identité numérique et ce phénomène s'explique grâce à la porosité.

« En 2009, les frontières de Facebook deviennent accessibles aux moteurs de recherches [...]. »¹²¹ Dès lors, nous avons assisté à un véritable éclatement des frontières sur le Web. Le modèle d'affaire des réseaux sociaux repose dorénavant sur la marchandisation des données personnelles et c'est, en grande partie, grâce aux techniques algorithmiques de ciblage publicitaire que cette stratégie est devenue économiquement profitable. Avec la marchandisation de ces fragments identitaires s'écroulent les frontières séparant jadis les multiples bases de données personnelles. À ce jour, chaque empreinte numérique qu'un individu laisse involontairement derrière lui est susceptible de traverser les frontières du Web afin de s'agglutiner aux autres particules granulaires de son identité numérique. L'effritement des frontières numériques signifie qu'une compagnie comme Google pourrait potentiellement regrouper la quasi-totalité des informations définissant l'identité numérique d'un individu, et ce, à travers toutes les plateformes connectées qu'il utilisera au cours de sa vie. La multiplication des réseaux sociaux permet effectivement de recueillir un éventail d'informations toujours plus large sur les individus et ce phénomène gagnera assurément en amplitude avec les générations subséquentes pour lesquelles cette interconnexion se généralisera. Nous avons précédemment soutenu que la socialisation impliquait le calcul du dévoilement identitaire relatif aux différents contextes auxquels il était destiné. Similairement, un individu dévoilera certainement différentes dimensions de son identité s'il se trouve sur une plateforme de rencontre, et d'autres s'il est sur un réseau professionnel. La socialisation dirigeant ce choix découle de la simple présomption qu'il existe forcément des frontières entre ces entités numériques comme il existe des frontières entre la chambre à coucher et le bureau. Néanmoins, la réalité dément cette supposition et la porosité des frontières numériques explique comment la construction des identités numériques se situe à l'intersection de l'ensemble de nos activités connectées.

¹²¹ *Ibid.*, p. 45.

L'idée de la percolation illustre comment, grâce à chaque granule identitaire parsemant le parcours digital d'un individu, est constituée son identité numérique. Considérées individuellement, les activités d'un individu sur le Web n'ont aucune signification, mais regroupées et organisées entre-elles, elles acquièrent un sens ainsi qu'une valeur. Les géants du Web ayant les moyens d'accéder aux multiples bases de données personnelles définissant nos identités numériques sont d'ailleurs « [...] les seuls à disposer d'une vue synoptique de l'ensemble de nos comportements connectés [...] »¹²². C'est grâce à ces informations privilégiées, dans une majorité de cas associés à des identités civiles véritables, que certaines entreprises parviennent à assurer leur profitabilité. L'instrumentalisation des identités numériques revêt plusieurs visages et la gestion algorithmique des données repousse toujours les frontières du possible. À titre d'exemple, l'incident de Cambridge Analytica, faisant directement référence à l'exploitation des données personnelles étant associées à des identités numériques, avait permis à cette entreprise de manipuler idéologiquement le contenu informationnel soumis à des millions de profils de Facebook de manière à influencer l'élection présidentielle américaine de 2016. De manière plus anodine, les données personnelles des individus peuvent potentiellement être utilisées dans une multitude d'opérations algorithmiques allant du ciblage publicitaire aux analyses de risques effectuées par les compagnies d'assurances. La granularité, la porosité ainsi que la percolation sont donc des principes permettant d'explicitier la constitution de l'identité numérique telle qu'elle existe aujourd'hui. Les données personnelles desquelles résultent fondamentalement ces constructions devraient alors être considérées à l'aune des mesures englobant les enjeux de protection de la vie privée en IA, mais cette question dépasse les considérations de ce mémoire. Notre objectif est plutôt de démontrer le lien entre l'identité numérique et l'identité sociale, et c'est grâce à la socialisation, ainsi qu'au phénomène de la performativité algorithmique que nous y parviendrons.

2.3 La performativité algorithmique

Jusqu'à maintenant, nous avons préparé le terrain pour la théorisation du phénomène de la performativité algorithmique. En ce sens, nous avons successivement

¹²² *Ibid.*, p. 48.

expliqué le fonctionnement algorithmique ainsi que l'avènement de l'IA, le concept de performativité ainsi que son application au genre et à la construction identitaire intersectionnelle, pour terminer avec la génération des identités numériques à partir du pullulement des données personnelles. Comme nous l'avons déjà mentionné au début de ce chapitre, le phénomène de la performativité algorithmique se décline en deux moments, celui de l'émergence d'une force performative résultant de l'accumulation et de la répétition des décisions algorithmiques, ainsi que celui de l'effet performatif des complexes algorithmiques sur la construction identitaire. Ces deux moments seront donc explicités au cours des prochains points.

2.3.1 Le concept d'émergence

Nous savons que les algorithmes sont présents dans nos vies de tous les jours, nous savons qu'ils prennent des décisions, et que ces dernières génèrent des conséquences sur nos existences. De plus, la seule affirmation que les décisions algorithmiques ont des effets sur nos vies, et potentiellement sur la construction identitaire, n'a également rien de très surprenant. Si nous sommes évidemment en accord avec cette vraisemblance, nous croyons néanmoins que la seule influence des algorithmes sur la construction identitaire est insuffisante pour rendre compte adéquatement de ce qui se produit véritablement. Comme mentionné à maintes reprises, nous estimons que les décisions algorithmiques ont un effet performatif sur la construction identitaire. Néanmoins, il serait faux d'affirmer que les décisions algorithmiques sont intrinsèquement formatives puisque ces dernières, considérées singulièrement, ne s'accomplissent généralement pas. Il serait cependant juste de soutenir que certains complexes algorithmiques deviennent formatifs lorsqu'ils participent activement à la construction performative de l'identité des individus. C'est grâce au concept de l'émergence que nous expliciterons ce devenir, ou plus précisément, cette transition entre les simples processus décisionnels automatisés et les complexes algorithmiques formatifs aux fondements de ce mémoire.

« Le tout est plus que la somme de ses parties. »

Cette vieille et célèbre maxime, dont la provenance ne semble pas faire consensus, illustre l'essence même du concept d'émergence. Même si l'idée générale précède certainement l'utilisation officielle du terme, c'est en 1875, dans *Problem of Life and Mind*, que Georges

Henry Lewes mentionne que des entités émergentes peuvent être le résultat de l'action d'entités plus fondamentales et peuvent être parfaitement nouvelles ou irréductibles par rapport à ces dernières¹²³. Notamment mobilisé afin de réconcilier le matérialisme à la manifestation de la vie ou même de la conscience, le concept d'émergence soutient que même si les atomes singuliers et matériels ne sont ni vivants et conscients, ils peuvent le devenir grâce à leur structure. En admettant que l'organisation opportune d'atomes matériels permette l'émergence de propriétés nouvelles, nous soutenons qu'une force performative émerge de certaines architectures spécifiques de complexes algorithmiques. Objectivement, nous affirmons donc qu'une décision algorithmique prise simplement, bien qu'elle puisse avoir des conséquences directes sur le monde, ne peut être performative puisque la performativité implique une répétition organisée ainsi qu'une temporalité prolongée de cette dernière. Si la performativité algorithmique n'est pas un phénomène isolé selon nous, c'est parce que les complexes algorithmiques, sur lesquels nous reviendrons au prochain point, sont précisément caractérisés par la répétition soutenue de leurs actions. En effet, les grands systèmes d'algorithmes accomplissant le microciblage publicitaire, la recommandation de contenu, ou même la recherche d'informations, représentent tous des exemples d'importantes répétitions d'actions orchestrées, et donc structurées, en vue d'objectifs. Bien que ces visées soient généralement d'ordres économiques, nous persistons à croire que leur orientation favorise la construction performative de l'identité. Avant d'explicitier ce point, nous délimiterons les complexes algorithmiques puisqu'ils expliquent la répétition ainsi que l'accumulation des décisions permettant la force performative des algorithmes.

2.3.2 Les complexes algorithmiques

Au premier chapitre, il avait été question des différentes versions de Netflix ainsi que de l'intelligibilité provoquée par la complexité d'une telle plateforme. Grâce aux propos de Bucher, nous avons également mentionné que si les complexes algorithmiques étaient inintelligibles, ce n'était pas parce qu'ils devaient être considérés comme des boîtes noires, mais plutôt parce qu'il s'agissait d'accomplissement en devenir. Si les complexes algorithmiques sont changeants et dynamiques, qu'ils sont des *eventfuls* pour reprendre

¹²³ LEWES, G. H. *Problem of Life and Mind*, The Riverside Press, Cambridge, 1875, p. 174.

l'expression de Bucher, c'est parce que ces applications nécessitent des interactions et des interrelations entre un nombre important d'algorithmes. Comme le démontre la citation suivante, Netflix n'est pas constitué d'un algorithme, mais plutôt une constellation de processus décisionnels automatisés.

« Algorithms, though, are not stand-alone boxes but always part of complex systems. They often operate as a collection of algorithms in what are, ultimately, networked systems. This is particularly true for the algorithms underlying platforms such as Facebook, Twitter, Netflix, or YouTube. These systems do not contain one algorithm but a collection of algorithms, working together to create a unified experience. As product owners at Netflix, Gomez-Uribe and Hunt write, “our recommender system consists of a variety of algorithms that collectively define the Netflix experience” (2015: 2). »¹²⁴

Lorsque nous utilisons la notion de complexe algorithmique, c'est à cette expérience unifiée que nous faisons référence, c'est-à-dire aux applications ainsi qu'aux plateformes que nous utilisons dans nos vies. Lorsque Netflix ou YouTube recommande du contenu, que Facebook ou Instagram suggère certaines publicités micro ciblées, et que Google prédit les résultats de recherches, il s'agit de décisions orchestrées par de vastes réseaux d'algorithmes interconnectés. Le terme *complexe algorithmique* saisit donc cette nuance qu'il faut mettre de l'avant entre un algorithme, et l'expérience unifiée émanant d'un système complexe d'algorithmes. Cette nuance est capitale puisqu'elle permet d'expliquer la répétition ainsi que l'accumulation des décisions générées par les plateformes et applications. Même si ces dernières diffèrent par ce qu'elles accomplissent, elles doivent essentiellement répondre à deux besoins, soit l'expérience agréable et cohérente des utilisateurs, ainsi que la rentabilité de leur modèle d'affaires. En ce sens, ces deux besoins seraient suffisants pour motiver la répétition unifiée des décisions algorithmiques générées, et cela expliquerait ultimement la tendance qu'on les complexes algorithmiques à générer une force performative. Si nous prenons l'exemple du microciblage publicitaire sur une plateforme comme Facebook, l'expérience agréable du fil d'actualité est effectivement propice à la suggestion répétée d'items pour les utilisateurs. Bien que Facebook ne soit pas un détaillant de marchandise, il n'en demeure pas moins que pour des fins de viabilité

¹²⁴ BUCHER, Tania. *Op. cit.*, p. 11.

économique, Facebook a su profiter des données personnelles générées par les différents profils afin de maximiser la précision du microciblage publicitaire sur sa plateforme. Si le ciblage est pertinent, l'expérience du client sera d'autant plus cohérente, et la stratégie publicitaire potentiellement plus lucrative. Si nous nous attardons à l'éclaircissement de ce fonctionnement, c'est pour démontrer que les complexes algorithmiques sont effectivement programmés pour que la répétition de leurs actions s'accumule, ce qui justifierait l'émergence de leur force performative. Si la répétition est une condition nécessaire à l'effet performatif des algorithmes sur la construction identitaire, elle n'est pas suffisante pour assurer sa manifestation. Nous soutenons effectivement qu'une autre condition de ce phénomène réside dans l'interaction performative entre les individus et les algorithmes, ce que nous expliquerons dans les lignes qui suivent.

Bien que le fonctionnement des complexes algorithmiques permette l'émergence d'une force performative, la profondeur de ces derniers occasionne un autre problème, celui de leur intelligibilité. Si l'analogie classique boîtes noires est selon nous stérile pour la compréhension des complexes algorithmiques, la proposition de Bucher, soutenant qu'il s'agit plutôt d'*eventful*, ou d'*ongoing accomplishment*, permet une certaine transparence de ces derniers. Malgré le fait que la permutation constante des complexes algorithmique obscurcisse leur compréhensibilité, la notion d'*eventful* ouvre la porte à une interaction performative entre les différentes plateformes/applications et les individus/utilisateurs. « Here, I suggest a conceptual vocabulary that allows for investigating algorithms as *eventful*, understood as constituents that co-become, and their power and politics as tied to the ways in which these configurations have the capacity to produce certain orderings and disorderings of the world. »¹²⁵ Les algorithmes, érigés en complexes, se transforment et sont constamment en devenir. Les décisions qu'ils génèrent ne seront pas toujours les mêmes, car elles sont influencées par une myriade de facteurs découlant des variations qu'ils subissent. L'apprentissage constant des algorithmes explique effectivement, du moins partiellement, ces variations, et pour cette raison, les données avec lesquelles ils apprennent orientent certainement leurs modulations. Ces données servent à l'apprentissage machine et peuvent avoir des sources extrêmement variées. Comme nous

¹²⁵ *Ibid.*, p.17.

le verrons dans le prochain chapitre, ces dernières soulèvent inévitablement la question des biais que peuvent reconduire les algorithmes. Avant d'en arriver à ces considérations, il est tout de même possible d'affirmer que les données personnelles occupent une place prépondérante dans l'entraînement des algorithmes, et ce, en raison des avantages concurrentiels que procure une meilleure connaissance des consommateurs/utilisateurs. Cette réalité génère certainement des modifications sur le fonctionnement des algorithmes étant donné que ces derniers sont considérés comme des accomplissements en devenir. En ce sens, nous pensons qu'il est également possible de soutenir que les individus ont un effet performatif sur ce que deviennent les algorithmes. Effectivement, puisque l'apprentissage machine a pour objet principal les activités humaines, alors il a en va de soi que ces dernières aient une influence sur la nature des décisions éventuellement prises. Si nous croyons que cette influence sera performative, c'est parce que la notion de performativité exige un va-et-vient, une interaction, entre les éléments se construisant simultanément.

2.3.3 Le renforcement performatif

Pour Butler, le genre performatif impliquait effectivement une interaction entre les individus genrés et le genre, entre la construction performative du genre et la construction performative des identités genrées, et c'est précisément ce va-et-vient performatif qui explique que le genre, tout comme l'identité genrée y étant associée, sont des constructions malléables et dynamiques, des accomplissements en devenir (*eventful*). Tout comme le genre a un effet performatif sur la construction identitaire à cause de la répétition et de l'accumulation des symboles hétéronormatifs que nous retrouvons dans nos sociétés, les complexes algorithmiques ont un effet performatif sur la construction identitaire des individus provoqué par la répétition et l'accumulation des décisions générées. Par contre, le genre, tout comme les complexes algorithmiques et la nature des décisions qu'ils génèrent, sont inversement et performativement construits par les actes et agissements, ou performances, des individus. Si les individus sont à la base du genre parce qu'ils le composent et l'incarnent, les individus sont à la base des décisions algorithmiques puisque leurs performances, ou les actes et agissements/comportements enregistrés en tant que données personnelles, constituent le matériau d'apprentissage de ces processus décisionnels automatisés.

Afin d'illustrer ce processus, il faut prendre l'individu, ainsi que ses actions et agissements, comme point de départ. Lorsque ces informations sont enregistrées et qu'elles deviennent des données personnelles, elles peuvent par la suite être utilisées pour l'apprentissage machine, et ce sera à partir de ce moment précis que les activités humaines auront un effet performatif sur les accomplissements en devenir que sont les algorithmes. Il est par exemple possible de penser au fonctionnement de Facebook ainsi qu'aux premières catégorisations de profils d'utilisateurs générées grâce à l'application des stratégies de clustering aux données personnelles. Effectivement, les règles de catégorisation des individus sont établies grâce aux informations fournies par ces mêmes individus. Poursuivant notre raisonnement, nous avons démontré que la répétition accumulée de décisions algorithmiques pouvait avoir un effet performatif sur la construction identitaire. Lorsque ces décisions sont orchestrées et coordonnées en fonction des catégorisations algorithmiques associées à un individu, nous croyons qu'il est juste d'affirmer que ces dernières auront effectivement un effet performatif sur l'identité de ce dernier. Plus l'identité numérique d'un individu sera détaillée et complexe, et plus les décisions algorithmiques prises à l'égard de cet individu seront personnalisées et synchronisées avec les préférences et champs d'intérêts de ce dernier. Grâce au nombre incommensurable de données et d'identités numériques qu'elle manie, une plateforme comme Facebook est capable d'anticiper, de prédire, et même de provoquer certaines préférences et champs d'intérêts grâce à l'apprentissage profond. En influençant à répétition le développement d'un individu grâce à des modèles prédictifs, nous jugeons qu'il devient possible de parler d'effet performatif des décisions algorithmiques sur la construction identitaire. Nous croyons néanmoins que le phénomène de performativité algorithmique s'illustre réellement dans sa tendance à s'autorenforcer. Effectivement, une fois que la construction identitaire est influencée par la catégorisation algorithmique, les agissements de ces mêmes individus seront ensuite récupérés, encore sous la forme de données personnelles, par les algorithmes de catégorisations et de prédictions. Suite à cela, les catégories (*cluster*) générées par ces algorithmes seront automatiquement plus représentatifs des individus à cause de ce va-et-vient performatif. Comme nous le verrons lors du prochain chapitre, ce processus de rétroaction en boucle (*feedback loop*) fait en sorte que certains biais stigmatisants peuvent s'infiltrer au sein de ce phénomène complexe.

Avant d'y arriver, nous allons présenter comment l'interaction performative entre les individus et les algorithmes peut concrètement constituer l'identité.

2.3.4 La socialisation de l'identité

Si nous récapitulons, la performativité algorithmique fait référence à l'effet performatif des algorithmes sur la construction identitaire des individus. La réinterprétation du concept de performativité de Butler soutient que le genre, tout comme l'identité genrée, sont coconstruits dans une interaction performative. C'est grâce à la répétition des actes, agissements et comportements que les individus constituent le genre, et inversement, ces derniers incarnent également le genre lors de leurs performances sociales. L'identité sociale d'un individu est performativement construite grâce à l'accumulation des performances sociales qu'il incarne. La même interaction performative existe dès lors entre les groupes sociaux, et les individus, et le tout s'articule dans un cadre intersectionnel. L'identité sociale d'un individu est ultimement déterminée à l'intersection des groupes sociaux auxquels il appartient, et pour cette raison, l'identité sociale d'un individu est beaucoup plus complexe que la simple addition des différentes appartenances de ce dernier. Les multiples intersections aux fondements d'une identité sociale sont hiérarchisées selon des rapports de forces et des contextes situationnels, et l'identité d'un individu est pour cette raison mouvante. À titre d'exemple, un homme blanc et catholique appartient et représente simultanément trois groupes sociaux, celui des hommes, celui des blancs ainsi que celui des catholiques. L'identité sociale de cet individu, se retrouvant à l'intersection de ces trois groupes sociaux hiérarchiquement positionnés, est une construction performative puisqu'elle est ultimement le fruit d'une performance sociale articulant organiquement ces trois réalités. Comme c'était le cas avec l'analogie de la métaphore théâtrale de Goffman, le contexte situationnel aura une incidence sur le positionnement de l'intersection identitaire et c'est pour cette raison que la performance sociale incarnée sera différente dans une église que dans une salle de sport. En contrepartie et comme Butler l'affirme, les individus deviennent les performances qu'ils incarnent, et c'est en ce sens que la construction identitaire sera, pour les besoins de ce mémoire, performative et intersectionnelle.

Cette conception de la construction identitaire a également le mérite de réconcilier le fonctionnement des algorithmes à la réalité sociale des individus. Effectivement, le concept

de groupe social de Young renvoie directement aux *clusters* générés par la catégorisation algorithmique des identités numériques. Si ces ensembles diffèrent selon plusieurs critères, il n'en demeure pas moins que dans les deux cas, ces derniers sont changeants et performativement construits. Si l'identité sociale est associée à l'appartenance intersectionnelle à des groupes sociaux, l'identité numérique est similairement associée à l'appartenance intersectionnelle à des catégorisations algorithmiques. Bien que la nature de l'identité sociale et de l'identité numérique soit intrinsèquement différente, l'interaction entre ces dernières est défendable. Effectivement, les processus de socialisation migrent vers un univers de plus en plus connecté ou les innovations, comme c'est le cas avec la transformation des réseaux sociaux vers le Métavers par exemple, transforment les interactions. Si l'effet du phénomène de performativité algorithmique sur la construction identitaire est moins flagrant lorsqu'il est question d'individus traditionnellement socialisés, il se manifeste avec plus d'évidence lorsque l'on considère la socialisation des générations actuelles ainsi que de celles qui viendront. Bien qu'il existe plusieurs définitions de la socialisation entre les approches psychologiques et sociologiques, il est prudent d'affirmer qu'une majorité d'entre elles théorisent le processus de « training or preparing an individual to fit into a given culture or social group. »¹²⁶ La socialisation s'exerce potentiellement sur un individu tout au long de sa vie sociale, mais son importance est clairement marquée lors des premiers stades de la vie, soit l'enfance, associée à la bulle familiale, ainsi que l'adolescence, associée à la sortie de cette cellule¹²⁷.

« Socialization has as main purpose the realization of the ability to obey the exigencies of social norms and ideals materialized in the cultural values of the group to which man belongs. The content of socialization has not only a psychological significance (the maturing of the young person), but also a culturological (internalizing social norms and values) and sociological (social role-playing and development of appropriate behaviors). [...] The socialization process represents the social integration and identity of the person. »¹²⁸

Avec l'avènement des réseaux sociaux, des jeux vidéo en ligne, et de toutes les plateformes et applications interconnectées, la socialisation des jeunes s'effectue de plus

¹²⁶ EAGLE, Gill, « Learning to Become a "Natural Woman": The Process of Socialisation », *Empowering Women for Gender Equity*, no 2, 1988, p. 67-80.

¹²⁷ PESCARU, Maria, *The importance of the socialization process for the integration of the child in the society*, University of Pitesti, 2019, p. 21.

¹²⁸ *Ibid.*, p.20.

en plus par l'entremise des nouvelles technologies et identités numériques y étant associées. Le contenu de cette socialisation, sur le plan culturel et social, participe activement à la construction de l'identité numérique, mais également à celle de l'identité sociale. Si la distinction entre ces deux identités peut paraître évidente aux yeux de ceux qui ont vu naître ces technologies, les nouvelles formes de socialisations remettent certainement en question cette frontière. La socialisation, en tant que processus au fondement de l'identité, explique donc l'interdépendance ainsi que l'interaction grandissante entre l'identité numérique et l'identité sociale. Les actes et agissements des individus, en tant qu'actions participant à la construction performative des groupes sociaux et générant simultanément la catégorisation algorithmique de ces derniers, résultent de la socialisation et de son contenu. Pour reprendre l'exemple de la construction performative du genre de Butler, les hommes et les femmes sont socialisés, dès leur plus jeune âge, pour se comporter, et donc incarner, une performance sociale masculine ou féminine. De nos jours, il ne fait aucun doute que cette socialisation, qu'elle soit genrée ou associée à un groupe social d'une autre nature, soit reconduite par la catégorisation algorithmique et son effet sur l'identité numérique.

Si nous prenons l'exemple de la socialisation d'un individu né dans les années 2000 dans un milieu industrialisé, nous sommes convaincus qu'il est juste d'affirmer que le phénomène de la performativité algorithmique aura participé à sa construction identitaire. Le cas de figure d'une jeune femme noire nous permettrait d'illustrer cette réalité. En venant au monde, cette dernière sera socialisée, au sein de sa propre famille, comme étant femme et comme étant noire, c'est-à-dire qu'elle comprendra graduellement son appartenance *naturelle/objective* à ces groupes sociaux distincts, son identité s'articulant progressivement à cette intersection. Bien entendu, l'acceptation de cette appartenance n'est pas infaillible et de nouvelles appartenances peuvent évidemment s'y amalgamer. Pour des soucis de clarté, nous nous limiterons néanmoins à ce cas simple. Au cours de sa socialisation, cette jeune femme prendra tranquillement conscience des limites de sa cellule familiale, et d'autres espaces de socialisation apparaîtront progressivement. L'école agit encore aujourd'hui comme un espace de socialisation facilitant le passage de la famille à la société. Si l'école et l'adolescence sont associées à une période charnière pour la construction identitaire des individus, c'est précisément à cause de cette fonction

socialisatrice. En contrepartie, cette période de transition a grandement été transformée par la multiplication des réseaux sociaux et autres plateformes/applications interconnectées. Si nous reprenons l'exemple de notre jeune femme noire, nous soutenons que le groupe social des femmes, ainsi que celui des noirs, sont des constructions performatives résultant de l'accumulation des actes et gestes de leurs représentants. En récoltant les données personnelles des femmes noires plus vieilles et précédant notre exemple, les algorithmes de catégorisation ont érigé, grâce à certaines ressemblances et différences au sein des données personnelles, des identités numériques appartenant à des *clusters*. De ces identités, les complexes algorithmiques ont appris à microcibler, à suggérer, à anticiper et à prédire les préférences/habitudes/tendances des individus y étant associés. Si l'action de la performativité algorithmique est moins prononcée pour les premières générations d'identités numériques, cette dernière a tendance à s'accroître avec l'accumulation des données personnelles et donc des générations. En se socialisant davantage sur des réseaux sociaux, la performance sociale de notre exemple est virtuellement reconduite, catégorisée, et associée aux *clusters* regroupant des identités virtuelles similaires. En prédisant une homogénéité des préférences/habitudes/tendances des individus appartenant aux mêmes *clusters*, les complexes algorithmiques génèrent des décisions, sous la forme de propositions/suggestions/prédictions, qui participeront et orienteront performativement la socialisation des jeunes. Si notre jeune femme noire est orientée vers un contenu de socialisation et d'interactions à cause de son appartenance à certains groupes sociaux, et que la nature de ce même contenu finit par définir les champs d'intérêt/croyances/affiliations de cette dernière, alors le ou les complexes algorithmiques n'auront pas seulement prédit ces préférences, mais les auront performativement construits. De plus, les algorithmes se renforceront dans cette boucle de rétroaction, et avec l'accumulation répétée des décisions algorithmiques générées, il sera éventuellement possible de percevoir les effets de la performativité algorithmique sur la construction identitaire. Cette même jeune femme noire, après s'être performativement construite à travers cette socialisation connectée et intelligente, laissera sur ses traces numériques des données personnelles qui viendront renforcer les mêmes décisions algorithmiques ayant activement défini son identité. Et c'est dans ce va-et-vient incessant qu'il est possible de

saisir l'interaction performative entre les individus et les algorithmes, ce qui explique l'action du phénomène de la performativité algorithmique sur la construction identitaire.

3 Normes de justice et biais algorithmiques

3.1 L'injustice algorithmique

Depuis qu'il est reconnu que les décisions algorithmiques affectent la vie des gens et qu'elles ont des conséquences sociales dans le sens le plus large, l'impératif moral et politique de justice est désormais une source de préoccupation importante dans le domaine de l'IA. Effectivement, informaticiens et programmeurs s'intéressent actuellement aux conceptions de justice et plusieurs aspirent à une définition mathématisable du juste. « *Most exciting fairness definitions and algorithms make the implicit assumption that an underlying procedural or statistical notion of fairness can be mathematically defined and operationalized to create a fair system.* »¹²⁹ La question de la mathématisation de la justice soulève l'enjeu du pluralisme des conceptions de justice s'appliquant potentiellement à la programmation algorithmique. Suresh et Guttag défendent l'idée que les critères de justice attribués aux différents algorithmes dépendraient dans les faits de leurs fonctions et objectifs.¹³⁰ Ainsi, dans le cadre d'une application médicale, il serait légitime que certaines caractéristiques comme celles du genre et/ou de l'ethnicité soient considérées étant donné qu'elles sont potentiellement importantes lors d'un diagnostic. En revanche, il serait inapproprié que ces mêmes critères soient pris en considération pour une application destinée à l'embauche étant donné que cela soulèverait des enjeux de discrimination. Un exemple comme celui-ci soulève la question des multiples normes de justice qu'il est possible d'appliquer aux décisions algorithmiques, lesquelles dépendent ultimement des enjeux en cause. S'il peut être considéré juste de permettre à tous l'accès à un diagnostic établi sur une base de données complète, il serait néanmoins considéré injuste que l'accès à l'emploi soit basé sur une base discriminatoire. La hiérarchisation des normes de justice varie donc selon la nature du problème traité et c'est pour cette raison que les modèles algorithmiques doivent être soumis individuellement à des normes de justice variées. Néanmoins, si l'on considère l'enjeu de notre mémoire, celui d'une construction identitaire sans interférence, nous croyons qu'une théorisation de la justice englobante est possible.

¹²⁹ SURESH, Harini et John V. GUTTAG, *A framework for understanding unintended consequences of Machine learning*, 2020, p. 4.

¹³⁰ *Ibid.*, p. 1-2.

Bien que l'injustice algorithmique puisse prendre plusieurs formes, nous nous concentrerons sur les injustices de représentation étant donné qu'elles sont susceptibles d'interférer avec la construction identitaire. Ce sujet est traité dans l'ouvrage phare de Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*. Ayant grandement inspiré l'objet de ce mémoire, les algorithmes de l'oppression, tels qu'ils sont décrits par Noble, déterminent comment le moteur de recherche de Google participe au renforcement du racisme. Elle écrit, « this book was born to highlight cases of such algorithmically driven data failures that are specific to people of color and women and to underscore the structural ways in that racism and sexism are fundamental to what I have coined *algorithmic oppression*. »¹³¹ Grâce à l'analyse de plusieurs résultats de recherches concernant une variété de profils (« “black girls,” “Latinas,” and “Asian girls” »), Noble démontre que l'algorithme PageRank de Google renforce et reconduit des stéréotypes racistes et sexistes en associant des recherches à des résultats de recherches stigmatisants. L'un des exemples les plus probants est celui de l'association entre la recherche “black girls” et le contenu pornographique les représentant.¹³² Comme le souligne Noble, le problème de ce résultat réside dans le pouvoir représentationnel du moteur de recherche de Google qui représente désormais une source d'information fiable pour une majorité. Effectivement, elle nous rappelle que Google répond à une logique commerciale, et que la sexualisation de certaines recherches est certainement lucrative considérant l'importance économique de l'industrie pornographique.¹³³ En associant des résultats pornographiques à la recherche “black girls”, PageRank renforce et reconduit structurellement des stéréotypes racistes et sexistes dont l'effet est assurément dégradant pour les individus représentés. Nous sommes en accord avec Noble pour affirmer que cette situation constitue un cas d'oppression et que cela constitue une injustice grave. Ajoutons toutefois que cette dernière ne suggère aucune norme de justice afin de prouver qu'il s'agit effectivement d'une injustice.

¹³¹ NOBLE, Safiya Umoja, 2018, *Op. cit.*, p. 17.

¹³² *Ibid.*, p. 112.

¹³³ *Ibid.*, p. 102.

3.2 Les conceptions de justice

3.2.1 La « *Théorie de la Justice* »

Pour établir qu'il y a une injustice algorithmique, nous devons d'abord mobiliser des principes de justice qui nous permettent de la reconnaître et de l'évaluer. Nous pouvons nous référer à la *Théorie de la justice* de John Rawls étant donné qu'il s'agit du socle fondateur des différentes conceptions contemporaines de justice sociale. L'objectif de cet ouvrage étant d'offrir une alternative à la doctrine utilitariste de Mill et de Bentham qui dominait tant la philosophie morale que politique, Rawls réhabilite la stratégie contractualiste, notamment celle de Locke, Rousseau et Kant, afin d'établir par un accord social les principes de la société juste. Pour arriver à cet accord, Rawls imagine une délibération entre des êtres rationnels et impartiaux, les « partenaires ». Afin de s'assurer de leur impartialité, ce dernier s'assure que leurs intérêts particuliers, étant motivés par leurs conditions matérielles et sociales, sont cachés derrière un voile d'ignorance. Cet outil conceptuel hypothétique permet donc la détermination et la justification des principes de justice rendant possible une société juste et équitable.¹³⁴ Effectivement, l'accord originel résultant de la délibération de ces êtres rationnels et impartiaux garantirait la légitimité des principes déterminés. Cependant, il est nécessaire de faire un retour sur la conception de la société de Rawls, car cette dernière est antérieure à l'accord originel. L'auteur présente la société comme une association dont les règles déterminent un système de coopération avantageant mutuellement ses membres grâce à la répartition des fruits de cette collaboration.¹³⁵ L'objet de la justice réside donc dans les institutions sociales puisque ces dernières établiront la juste distribution des bénéfices de la coopération sociale aux membres de la société,¹³⁶ c'est-à-dire dans la structure de base de la société. « *Pour nous, l'objet premier de la justice, c'est la structure de base de la société ou, plus exactement, la façon dont les institutions sociales les plus importantes répartissent les droits et les devoirs fondamentaux et déterminent la répartition des avantages tirés de la coopération sociale.* »¹³⁷ Si la conception de la justice de Rawls s'inscrit dans un paradigme distributif, c'est précisément parce que les principes de justices déterminés par l'accord originel régissent

¹³⁴ RAWLS, John, *Théorie de la justice* (traduit de l'anglais par Catherine Audard), 1971, p. 37-38.

¹³⁵ *Ibid.*, p. 30.

¹³⁶ *Ibid.*, p. 30.

¹³⁷ *Ibid.*, p. 33.

le fonctionnement de la structure de base, laquelle distribue et répartit le fruit de la collaboration sociale.

En ignorant leurs conditions matérielles et sociales, les individus derrière le voile de l'ignorance auront rationnellement avantage à déterminer deux principes de justice, le premier ayant préséance sur le deuxième grâce à une hiérarchie lexicale. Le premier principe de justice stipule que « chaque personne doit avoir un droit égal au système total le plus étendu de libertés de base égales pour tous, compatible avec un même système pour tous. »¹³⁸ Sommairement, ce principe soutient que la liberté doit être distribuée également et maximale à tous, sans que la liberté des uns n'empiète sur celle des autres. Le deuxième stipule que les inégalités économiques et sociales sont justifiées pourvu qu'elles participent « au plus grand bénéfice des plus désavantagés »¹³⁹, il s'agit du principe de différence. Toutes réductions des libertés doivent donc favoriser le système de liberté partagé par tous, tandis que les inégalités doivent être profitables pour tous, notamment grâce à la redistribution des structures de base.¹⁴⁰ Même s'il était possible à ce stade d'imaginer l'articulation de ces principes autour des injustices algorithmiques, nous jugeons qu'il est préférable d'approfondir la pensée rawlsienne puisque ses travaux ultérieurs faciliteront cet exercice. Effectivement, ces principes constituent les contours d'une pensée libérale au sein de laquelle une place centrale est accordée au respect des libertés individuelles, tandis que la fonction de la structure de base de la société juste est de neutraliser les déterminismes naturels et sociaux.¹⁴¹ Rawls publiera *Libéralisme politique* en 1995, ouvrage permettant une compréhension concrète de l'application politique de ces principes.

3.2.1.1 Le « Libéralisme politique »

Articulé autour de la neutralité, le libéralisme politique de Rawls est fondé sur l'idée que l'État doit être neutre face aux conceptions de la vie bonne et de la personne, c'est-à-dire que les individus doivent avoir la liberté d'adhérer aux croyances/idéaux/valeurs de leur choix sans que les institutions sociétales ne s'immiscent dans ces décisions. Dans

¹³⁸ *Ibid.*, p. 341.

¹³⁹ *Ibid.*

¹⁴⁰ *Ibid.*

¹⁴¹ DESMONS, Ophélie, *Les présupposés du libéralisme politique : quelle justification?*, 2014, p. 22.

Théorie de la justice, Rawls s'oppose effectivement au perfectionnisme¹⁴², doctrine éthique et politique admettant que les institutions doivent orienter le perfectionnement de l'être humain vers l'excellence, cette dernière impliquant nécessairement une conception de l'homme et de la vie bonne vers laquelle tous devraient tendre. Là où le paternalisme admet qu'il est légitime de promouvoir ce qui est bon pour la société, le libéralisme s'interdit cette prétention de par sa neutralité.¹⁴³ Effectivement, l'État libéral doit demeurer neutre puisqu'étant soumis au pluralisme des conceptions de biens auxquelles les citoyens adhéreront, il deviendrait perfectionniste si ses institutions sociales favorisaient certaines d'entre elles. Rawls fait ainsi référence au pluralisme raisonnable en reconnaissant qu'une « société démocratique moderne est caractérisée, non seulement par une pluralité de doctrines compréhensives, morales, philosophiques et religieuses, mais aussi par le fait que ces doctrines sont incompatibles entre elles tout en étant raisonnables. »¹⁴⁴ Si les citoyens sont définis comme libres et égaux dans un contexte démocratique, ils doivent également être rationnels et raisonnables, dans le sens qu'il serait déraisonnable d'utiliser le pouvoir politique afin d'imposer une conception du bien.¹⁴⁵ Cette définition négative de la neutralité politique n'étant toujours pas satisfaisante pour nos considérations, nous explorerons la proposition positive de Rawls à l'égard de cette notion.

3.2.1.2 *Le concept de neutralité*

La neutralité devient donc la norme à respecter des politiques libérales. Rawls souligne que les exigences de la neutralité « ne s'appliquent pas à toutes les questions politiques », mais « s'appliquent seulement à celles qui impliquent ce que nous avons appelé des questions constitutionnelles essentielles et des questions de justices fondamentales »¹⁴⁶, lesquelles ont le pouvoir d'avoir un impact sur la justice sociale. À partir du moment où la neutralité politique s'applique aux décisions politiques, trois sens peuvent être déclinés de cette dernière, la neutralité des conséquences, des buts, et des justifications. Dans un premier lieu, la neutralité suppose que les actions de l'État sont neutres relativement aux conséquences de ces dernières, et donc qu'il ne favorise pas la

¹⁴² RAWLS, John, (1971) *Op. cit.*, p. 363.

¹⁴³ DWORKIN, Ronald. *Op. cit.*, p. 51.

¹⁴⁴ RAWLS, John, *Libéralisme politique* (traduite de l'anglais par Catherine Audard), 1995, p. 4.

¹⁴⁵ *Ibid.*, p. 90.

¹⁴⁶ *Ibid.*, p. 261.

réalisation de certaines conceptions du bien au détriment d'autres. Le problème avec cette conception de la neutralité, c'est que toutes actions de l'État entraînent inévitablement des conséquences, et que puisqu'il est impossible de calculer l'effet réel et singulier de chacune de ces influences à l'échelle d'une population, il ne serait pas réaliste ni souhaitable d'exiger que ces dernières soient neutres. En rejetant la neutralité des conséquences, il propose, en second lieu, la neutralité des buts qui stipule que « l'État ne doit rien faire pour favoriser ou promouvoir une doctrine compréhensive particulière plutôt qu'une autre ou fournir davantage d'assistance à ceux qui en sont partisans »¹⁴⁷. Il s'agit d'une neutralité des intentions dans le sens où les intentions du législateur se doivent être neutres. Même si l'écueil de la neutralité des conséquences est dépassé, de nouveaux problèmes émergent comme celui de la qualification de l'intention derrière les décisions collectives, ou encore de la connaissance évidemment incertaine des intentions d'un individu. Pour ces raisons et dans un troisième lieu, Rawls propose finalement la neutralité des justifications signifiant que l'État doit justifier la conception de la justice selon une procédure ne faisant intervenir aucune conception compréhensive du bien. Étant donné qu'une justification implique nécessairement la mobilisation de certains présupposés substantiels susceptibles de faire la promotion d'une conception de la vie bonne, la justification neutre devra être basée sur des « idées politiques du bien » qui découlent de la « raison publique »¹⁴⁸.

« Nous devons distinguer entre, d'une part, une base publique de justification qui soit acceptable d'une manière générale par tous les citoyens – en ce qui concerne les questions politiques fondamentales – et d'autre part les nombreuses sources non publiques de justification qui font partie des doctrines compréhensives et ne sont acceptables que pour ceux qui adhèrent à celles-ci. [...] Les idées de bien dans la conception politique doivent être politiques et donc distinctes des idées du bien propres à des doctrines plus larges. »¹⁴⁹

C'est donc la base publique de justification des idées du bien qui permet de les rattacher à ce que Rawls appellera « la raison publique » qu'il faut distinguer des « raisons privées », dans le sens que cette base repose « sur des vérités simples, maintenant largement acceptées ou disponibles pour les citoyens d'une manière générale »¹⁵⁰. Effectivement, là où la raison

¹⁴⁷ *Ibid.*, p. 236.

¹⁴⁸ *Ibid.*, p. 259.

¹⁴⁹ *Ibid.*, p. 7.

¹⁵⁰ *Ibid.*, p. 272-273.

publique est simple et générale, qu'elle provient d'un savoir partagé par la société tout entière, les raisons privées sont complexes et particulières puisqu'elles découlent des savoirs propres à des associations socialement situées¹⁵¹. Selon Rawls, la raison publique n'est pas controversée puisqu'elle fait référence aux « idées politiques du bien », c'est-à-dire aux conceptions du bien qui ne s'appliquent qu'à la structure de base de la société, ou plus précisément aux institutions déterminant la juste distribution des fruits de la collaboration sociale. Il va donc de soi que l'État libéral n'a pas à être neutre à l'égard des « idées politiques du bien » étant donné qu'elles existent en vertu de la redistribution sociale qu'elles structurent.

3.2.1.3 *L'analyse des enjeux algorithmiques*

Comme nous venons de l'indiquer, le libéralisme politique de Rawls n'exige pas la neutralité de l'entièreté des décisions politiques, mais plutôt de celles s'appliquant à des questions constitutionnelles essentielles ainsi qu'à des questions de justice fondamentale. Si nous souhaitons détecter les injustices algorithmiques grâce à cette conception de justice, il serait essentiel d'accorder un statut particulier aux décisions algorithmiques puisqu'il serait parfaitement acceptable que les algorithmes privés participent à la promotion de certaines conceptions de la vie bonne étant donné que les intérêts privés ne sont pas responsables de la neutralité de l'État, et ne sont donc pas soumis à ses exigences.

En revanche, même si nous revenons aux exigences de neutralité auxquelles devraient être soumis les algorithmes institutionnels, il devient d'autant plus délicat de déterminer où se situe l'injustice des décisions algorithmiques générées. Selon la neutralité des justifications qui constitue la seule neutralité qu'il est nécessaire de préserver selon Rawls, cela implique que les décisions algorithmiques injustes seraient celles dont le but, et plus précisément la justification de ce but, est de favoriser une conception de la vie bonne. Lorsque l'on considère les algorithmes apprenants dont nous avons cité le fonctionnement au premier chapitre, il faut comprendre que leur programmation est orientée en vue de la résolution d'un problème, et non de la promotion d'un idéal quel qu'il soit. Il devient donc très difficile de justifier les valeurs qu'un algorithme promeut étant donné qu'il n'est pas programmé à cette fin, ces dernières sont plutôt collatérales. Bien

¹⁵¹ *Ibid.*

entendu, il serait possible grâce au cadre théorique de Rawls d'affirmer que des algorithmes programmés dans le but de discriminer sont injustes, mais il serait par exemple impossible de soutenir que l'algorithme PageRank de Google est injuste. Étant donné que la programmation de cet algorithme a pour but et justification de diriger les recherches des utilisateurs vers les résultats de recherches qu'ils sont le plus susceptibles de rechercher, alors les résultats de recherches générés ne peuvent pas être considérés comme injustes puisqu'ils n'ont pas été programmés en ce sens. Comme le démontre d'ailleurs Rawls, toutes les actions politiques auront forcément des conséquences qui avantageront et désavantageront certains individus, et l'exigence de neutralité de ces actions ne peut pas être formulée par rapport aux conséquences qu'elles engendrent. Confrontés à l'exemple de l'oppression des algorithmes de Noble, nous n'aurions d'autre choix que d'accepter que les décisions algorithmiques générées à l'égard des femmes noires constituent seulement les conséquences d'un algorithme initialement neutre. Puisque la neutralité des conséquences est invalidée par Rawls, nombre de décisions algorithmiques potentiellement injustes passent sous le radar, et c'est notamment le cas de la sexualisation des femmes noires généré par PageRank.

En considérant l'injustice algorithmique par rapport au cadre neutraliste de Rawls, nous concluons donc que cette conception de la justice est insuffisante pour traiter des enjeux de justice soulevés par le déploiement de l'IA. Effectivement, la distinction entre les exigences de neutralité des sphères publiques et privées engendre certains problèmes que nous pourrions possiblement surmonter, mais la disqualification de la neutralité des conséquences exclut néanmoins la quasi-totalité des injustices de représentation provoquées par les algorithmes. Étant donné que l'objet de ce mémoire est d'avoir une perspective globale de l'effet des décisions algorithmiques sur la justice sociale, il est impératif que la conception de justice retenue permette le traitement de l'entièreté des situations d'injustice engendrées par les processus décisionnels automatisés.

3.2.2 La Justice et la « *politique de la différence* »

3.2.2.1 *La distinction entre théorie idéale et non-idéale*

Dès le tout début de son ouvrage majeur, Young affirme d'emblée qu'elle ne construit pas une théorie de la justice¹⁵², mais qu'elle s'inscrit plutôt dans une perspective critique à l'égard de cette dernière. Héritier de l'école de Francfort, le courant de la théorie critique développe est basé sur l'émancipation, et s'oppose avec le postulat qu'une théorie peut se dissocier du contexte social et historique afin d'avoir une prétention à l'universel. Cette posture refuse la séparation entre le théoricien et l'objet, et rejette donc la dissociation entre la théorie et la pratique. En ce sens, la démarche de Young consiste à proposer une solution pratique à une réalité existante. Young qualifie la théorie critique de réflexions normatives historiquement et socialement situées, lesquelles prennent assises dans l'expérience sociale afin d'orienter leur pouvoir émancipateur¹⁵³. Pour cette raison, elle écrit : « I argue that instead of focusing on distribution, a conception a justice should begin with the concept of domination and oppression »¹⁵⁴, ces derniers concepts étant caractéristiques de la réalité sociale, ou plus précisément de l'expérience réelle de situations de domination et d'oppression vécues par plusieurs groupes sociaux aux États-Unis dans les années 1960 et 1970¹⁵⁵. C'est donc grâce à une analyse critique, d'une situation sociale donnée, que Young aspire à l'édification de la conception pratique de la justice.

D'un point de vue méthodologique, il est possible de faire une distinction entre la *Théorie de la justice* rawlsienne, et la conception de la justice comme *Politique de la différence* de Young, en termes de théorie idéale et non idéale¹⁵⁶. Comme Pierre-Yves Néron le présente, « ideal theorizing tries to identify and justify the principle of justice that individuals and institutions should follow in a society that is already just, while non-ideal theorizing considers the demands of justice for a world like ours, in which there are systematic and historic sources of ongoing injustice »¹⁵⁷. Cette nuance caractérise ce que tentent d'établir ces deux théories. Comme l'indique lui-même Rawls, la « position

¹⁵² YOUNG, Iris Marion, *Justice dans the politics of difference*, 1990, p. 3.

¹⁵³ *Ibid.*, p. 5.

¹⁵⁴ *Ibid.*, p. 3.

¹⁵⁵ *Ibid.*, p. 7.

¹⁵⁶ NÉRON, Pierre-Yves, *Rethinking the very idea of egalitarian markets and corporation*, 2015, p. 99.

¹⁵⁷ *Ibid.*

originelle n'est pas conçue comme étant une situation historique réelle, encore moins une forme primitive de la culture. Il faut la comprendre comme étant une situation purement hypothétique, définie de manière à conduire une certaine conception de la justice »¹⁵⁸. En ce sens, les principes de justice rawlsiens découlent effectivement de la construction d'une situation idéale au sein de laquelle la délibération des partenaires est régie par des valeurs dont la justesse est toutefois établie pas le théoricien lui-même. Nous pouvons donc affirmer que la *Théorie de la justice* appartient aux théories idéales tandis que la posture critique de Young, et par extension sa conception de la justice, se rattache aux théories non idéales. Ayant comme point de départ l'oppression et la domination, la *Politique de la différence* considère que l'objectif de la justice n'est pas d'ériger des principes généraux applicables en tout temps, mais plutôt d'analyser les situations d'injustices afin de définir les principes de justice nécessaires pour faire cesser l'injustice. Considérant l'incapacité de la théorie rawlsienne à détecter les injustices algorithmiques, nous démontrerons que ces dernières constituent des situations d'oppression et qu'il est d'ailleurs théoriquement possible d'y mettre terme.

3.2.2.2 *Les groupes sociaux*

Avant de définir l'oppression telle qu'elle est théorisée chez Young, la notion de groupe social doit être préalablement approfondie. Dans ce cadre théorique, le groupe est fondamental puisqu'il est socialement antérieur à l'individu. « While groups do not exist apart from individuals, they are socially prior to individuals, because people's identities are partly constituted by their group affinities. Social groups neglect ways that people identify themselves to others, which lead them to associate with some people more than others and to treat others as different. Groups are identified in relations to one another. Their existence is fluid and often shifting, but nevertheless real »¹⁵⁹. Il faut comprendre que la priorité sociale du groupe s'explique par la construction identitaire, car la genèse de l'individualité réside dans l'appartenance à un groupe social. Les injustices structurelles, telles que la domination et l'oppression, existent et sont vécues à cause des différences entre les groupes sociaux. Young écrit en ce sens : « I argue that where social group difference exists and some groups are privileged while others are oppressed, social justice

¹⁵⁸ RAWLS, John, (1971) *Op. cit.*, p. 38.

¹⁵⁹ YOUNG, Iris M., (1990) *Op. cit.*, p. 9.

requires explicitly acknowledging and attending to those group difference to undermine oppression. »¹⁶⁰. La politique de la différence fait directement allusion à cette différence de groupe, et si le concept de groupe social est central chez Young, c'est parce qu'il permet ultimement une meilleure compréhension de la dimension structurelle des injustices.

Considérant la performativité des algorithmes, il est intéressant de constater que les stratégies algorithmiques de *clustering* génèrent des catégorisations d'individus en fonction de leurs caractéristiques partagées, et que ces groupes sociaux artificiellement construits participent à l'interaction performative à la base de la construction identitaire. Il faut également noter que les groupes sociaux de Young sont des représentations dont le dynamisme intrinsèque implique une très grande fluidité et la nature définitionnelle de ces différents groupes est donc appelée à être modifiée à travers le temps et l'espace. Même si les groupes sociaux observables dans la société diffèrent initialement des *clusters* générés artificiellement, nous croyons que cette divergence tend à se minimiser à cause du phénomène de la performativité algorithmique. Puisque les décisions algorithmiques ont un effet performatif sur la construction identitaire, nous pouvons supposer que les caractéristiques identitaires des individus constituant les groupes sociaux réels se transformeront et s'agenceront ultimement avec les caractéristiques générées par l'action de la catégorisation algorithmique. À partir du moment où les algorithmes ont un effet performatif sur la construction identitaire en participant performativement à la délimitation des groupes sociaux, ils soulèvent alors un enjeu de justice représentationnelle.

3.2.2.3 *Les injustices structurelles et représentationnelles*

Lorsque nous avons affirmé que la conception rawlsienne de la justice était insuffisante pour détecter les injustices algorithmiques, c'était en partie parce qu'elle n'admettait pas l'existence des injustices représentationnelles qui sont caractéristiques du phénomène de la performativité algorithmique. Même s'il existe des injustices structurelles chez Rawls, ces dernières apparaissent comme des effets structurels puisqu'elles résultent de la structure de base, et ultimement, de la gestion des fruits de la collaboration sociale effectuée par les institutions. Elles découlent essentiellement d'une distribution inéquitable des biens matériels et immatériels issue de la coopération, et elles sont donc de nature

¹⁶⁰ *Ibid.*, p.3.

distributive. En contrepartie, les injustices structurelles chez Young concernent les structures relationnelles de pouvoir unissant les groupes sociaux entre eux. La distinction à faire entre ces deux conceptions de l'injustice structurelle réside donc dans la nature de ces dernières puisque chez Rawls, elle est institutionnelle, tandis que chez Young, elle est relationnelle, représentationnelle et symbolique. Dans les faits, la théorie rawlsienne ne s'intéresse pas à la représentation qu'on les groupes sociaux d'eux-mêmes, représentations qui existe à travers le prisme de l'impérialisme culturel chez, notion sur laquelle nous reviendrons plus en détail. Préalablement, nous aborderons l'enjeu de la normativité, ainsi que la question de la représentation y étant rattachée, puisqu'ils sont intimement liés à cette dernière.

La normativité est intrinsèquement liée à la notion de représentation puisque ce sont les groupes sociaux dominants, agents de l'impérialisme culturel, qui déterminent, dans un jeu interactionnel parfois involontaire, la frontière entre la normalité et la déviance, entre le supérieur et l'inférieur.¹⁶¹ La réalité sociale des membres d'un groupe social est relative aux rapports relationnels qu'ils entretiennent face aux membres des autres groupes sociaux avec lesquels ils interagissent. Considérant la division du travail, les hommes bénéficient de l'impérialisme culturel masculin puisque traditionnellement, la représentation de leur groupe était associée au travail salarié tandis que la représentation du groupe des femmes était associée au travail domestique. De plus, l'organisation du pouvoir s'est érigée dans un système patriarcal et plusieurs mécanismes sociaux, par exemple la division genrée des tâches et du travail, renforcent cette réalité. La réalité sociale des femmes est donc normalisée dans ce rapport relationnel de subordination à cet impérialisme culturel. En stipulant que tous s'inscrivent socialement en fonction de la représentation sociale que leur confèrent leurs appartenances à des groupes sociaux, nous n'avons d'autre choix que d'accepter que pour chaque groupe subissant l'effet d'une injustice représentationnelle, il existe forcément un autre group en relation avec ce dernier profitant des fruits de cette dernière. C'est la nature de la relation unissant les différents groupes sociaux qui permet l'établissement, ainsi que la disposition, de structures occasionnant les cas d'injustices structurelles. Si l'injustice structurelle chez Young est plus à même de rendre compte des

¹⁶¹ *Ibid.*, p. 60.

enjeux soulevés par le phénomène de la performativité algorithmique, c'est précisément parce qu'elle se concentre sur la nature relationnelle des injustices de représentations. Ces dernières sont indissociables des effets performatifs de la catégorisation algorithmiques sur la construction identitaire des individus. Avant d'aborder cette question, nous concluons notre argumentaire justifiant notre positionnement théorique en élaborant les notions fondamentales de l'oppression et de la domination.

3.2.2.4 *L'oppression et la domination*

Si les concepts d'oppression et de domination n'apparaissent pas dans le lexique de la *Théorie de la justice*, ils deviennent fondateurs chez Young puisque leur déclinaison explique toutes les formes d'injustices structurelles. Lorsque Young affirme que l'oppression et la domination constituent le point de départ de sa conception de la justice, c'est parce que ces injustices constituent des freins à la réalisation de deux valeurs générales recherchées par la *politique de la différence*, c'est-à-dire : « 1) developing and exercising one's capacities and expressing one's experience, and 2) participating in determining one's action and the conditions of one's actions. [...] To these two general values correspond two social conditions that define injustice: oppression, the institutional constraint on self-development, and domination, the institutional constraint on self-determination. »¹⁶² Afin de clarifier la distinction entre ces deux notions, il faut comprendre que l'oppression, sous ses cinq facettes que nous déclinons, implique habituellement l'action de la domination, tandis que la domination peut exister hors des limites de l'oppression. Puisque la domination implique l'assujettissement à une structure décisionnelle hiérarchique, tous les individus peuvent se retrouver, à un moment ou à un autre, dans une situation de domination. Ceci, étant donné que la société, et plus précisément le marché du travail capitaliste et bureaucratique, sont érigés hiérarchiquement. La domination aura donc inévitablement un effet déterminant sur les actions prises dans un rapport hiérarchisé. Ce qu'il faut comprendre à ce point, c'est que des individus non opprimés, et n'appartenant donc pas à des groupes sociaux correspondant minimalement à l'un des cinq critères de l'oppression, peuvent être dominés dans les structures capitalistes et bureaucratiques qu'ils intègrent souvent au travail. Cette

¹⁶² *Ibid.*, p. 37.

distinction maintenant clarifiée, nous nous intéresserons désormais aux différences facettes de l'oppression, et plus précisément, à celle de l'impérialisme culturel.

3.2.2.5 *Le prisme de l'oppression*

D'emblée, l'oppression est une injustice structurelle puisqu'elle représente la situation d'un individu appartenant à un groupe défavorisé, et constitue un frein au développement personnel. Comme l'indique Young, l'oppression n'est pas forcément consciente puisqu'elle s'institutionnalise à travers les relations entre les différents groupes sociaux :

«In this extend structural sense, oppression refers to the vast and deep injustices, some groups suffer as a consequence of often unconscious assumptions and reactions of well-meaning people in ordinary interactions, media and cultural stereotypes, and structural features of bureaucratic hierarchies and market mechanisms – in short, the normal processes of everyday life. We cannot eliminate this structural oppression by getting rid of the rulers or making some new laws, because oppressions are systematically reproduced in major economic, political and cultural institutions. »¹⁶³

Les actions s'inscrivant dans la *vie de tous les jours* participent effectivement à la reconduction des situations d'oppression puisque les identités des individus sont imbriquées dans les groupes auxquels ils appartiennent, et que tous les groupes sociaux existent dans des relations de pouvoir les subordonnant¹⁶⁴. L'identité se construit selon une logique d'appartenance aux groupes sociaux, et dans le même sens, les groupes sociaux sont définis par ce sentiment identitaire partagé entre les membres d'un groupe. Le statut social, l'histoire commune, les formes culturelles ainsi que l'univers symbolique participent tous à des sentiments d'affinité entre les membres d'un groupe, lesquels se conjuguent dans la construction identitaire¹⁶⁵. Si l'appartenance à un groupe opprimé constitue une situation d'injustice, c'est parce qu'elle a un impact négatif sur la représentation qu'ont les individus d'eux-mêmes. Suivant ce même raisonnement, l'appartenance à un groupe profitant de l'impérialisme culturel a un impact positif sur la représentation que se font d'eux-mêmes les individus appartenant à ce groupe privilégié. La nature négative et positive de ces répercussions représente donc, conformément à la

¹⁶³ *Ibid.*, p. 41.

¹⁶⁴ *Ibid.*

¹⁶⁵ *Ibid.*, p. 44.

définition de l'oppression de Young, un soutien ou une entrave institutionnelle au développement personnel. Si la différenciation de groupe est inévitable pour nos sociétés, elle n'est pas oppressive en elle-même puisque c'est plutôt la représentation stigmatisante de ces derniers qui la rend comme telle. Young ira même jusqu'à affirmer que la différenciation de groupe est souhaitable puisque l'uniformisation de ces derniers impliquerait inévitablement le renforcement du pouvoir normatif de l'impérialisme culturel. La solution résidant donc dans la suppression des situations d'oppression, nous survolerons ses cinq catégorisations afin d'en établir les nuances.

Dans un premier lieu, l'exploitation est représentée en tant que processus de transfert des fruits du travail d'un groupe social vers un autre. S'inspirant de la théorie marxiste, la conception de Young admet la relation structurelle de l'exploitation entre les différents groupes en affirmant qu'elle dépasse la simple distribution inéquitable des biens produits. À titre d'exemple, les enjeux engendrés par la division genrée du travail ne se limitent pas à des conditions matérielles moindres pour les femmes, mais facilitent également le développement personnel des hommes. L'institutionnalisation des rapports relationnels asymétriques entre les groupes fait en sorte que les groupes opprimés, en subissant l'exploitation, participent à l'augmentation des conditions de possibilité pour les groupes privilégiés. En deuxième lieu, l'impuissance (*powerlessness*) est caractéristique d'une exploitation du statut professionnel des travailleurs se traduisant par un manque d'autorité et de respectabilité des individus appartenant aux groupes minoritaires. En troisième lieu, la marginalisation désigne l'impossibilité structurelle, pour certains groupes de participer à la vie sociale et politique, ce qui leur permettrait normalement de faire valoir leurs revendications de groupe. Cette dimension de l'exploitation indique donc l'existence de processus sociaux empêchant la juste participation des individus à la détermination ainsi qu'à la poursuite de leurs intérêts. En quatrième lieu, la violence en tant que dimension de l'oppression ne représente pas des actes violents comme tels, mais plutôt les contextes sociaux rendant possible, voire acceptable, leur prolifération. En ce sens, la violence devient systémique, et donc oppressive, à partir du moment où des circonstances relationnelles font en sorte que les membres d'un groupe social sont visés par son action. Dans certains cas, cette violence peut être tolérée, légitimisée et même institutionnalisée dépendamment des situations.

En dernier lieu, l'impérialisme culturel explique comment la culture dominante, tel qu'elle est véhiculée par la perspective et l'expérience d'un groupe social, s'universalise socialement pour s'imposer comme la norme sociale. L'impérialisme culturel renforce donc la représentation des groupes dominants en imposant la normalité, et même l'universalité, de leur univers symbolique. Ainsi, son pouvoir normatif oppresse inévitablement les autres groupes en déclarant les symboles de leur culture comme différents, inférieurs, voire déviants. Ce jeu interactionnel entre la normalité et la déviance engendre donc des injustices relationnelles de représentations symboliques puisqu'en pénalisant les groupes *différents* et en avantageant les groupes *normaux*, l'impérialisme culturel se renforce. Faisant référence à l'hégémonie normative de la réalité culturelle d'un groupe, le ou les groupes sociaux détenant ces monopoles représentationnels jouissent des avantages engendrés par l'oppression des groupes y étant soumis.¹⁶⁶ Bien que les cinq faces de l'oppression soient habituellement en interaction, chacune d'entre elle est suffisante pour déterminer l'oppression d'une situation. Young souligne néanmoins que la quasi-totalité des injustices structurelles implique l'oppression de l'impérialisme culturel, et pour cette raison, nous nous concentrerons dorénavant sur l'enjeu de la culture. Ce choix est justifié par la relation entre la culture et le pouvoir normatif de l'impérialisme culturel qui est certainement pertinente pour la question de la construction identitaire.

« Culture is the most general of the three categories of nondistributive issues I focus on. It includes the symbols, images, meanings, habitual compartments, stories, and so on through which people express their experience and communicate with one another. Culture is ubiquitous, but nevertheless deserves distinct consideration in discussions of social justice. The symbolic meanings that people attach to other kinds of people and to actions, gestures, or institutions often significantly affect the social standing of persons and their opportunities. In Chapters 2, 4, 5 and 6, I explore the injustice of cultural imperialism which marks and stereotypes some groups at the same time it silences their self-expression. »¹⁶⁷

Dans l'extrait qui précède, il est possible de constater que Young, tout comme nous le ferons, axe son analyse critique sur la prédominance de l'impérialisme culturel étant donné sa portée générale. Puisque le paradigme distributif est insuffisant pour rendre compte des enjeux soulevés par le déploiement de l'IA, nous sommes d'avis que l'étude des

¹⁶⁶ *Ibid.*, p. 59.

¹⁶⁷ *Ibid.*, p. 23.

situations d'injustices algorithmiques doit être réfléchi à travers le prisme plus large de de la culture en tant que catégorie non distributive. En ce sens, l'analyse de l'effet performatif des algorithmes sur la construction identitaire nous permet de comprendre et de mesurer le pouvoir symbolique et représentationnel de ces derniers grâce au concept d'impérialisme culture.

3.2.2.6 L'idéal de neutralité et l'impérialisme culturel

Un autre argument en faveur de la sélection de la conception de la justice de Young, outre le fait qu'elle permet une meilleure détection et compréhension des injustices algorithmiques de nature représentationnelle, réside dans le fait que l'application de neutralité rawlsienne participe au renforcement de l'impérialisme culturel. Afin d'expliquer cette affirmation, il est possible de contraster le voile de l'ignorance rawlsien à la célébration de la différence chez Young. D'une première part, les principes de justice déterminés derrière le voile de l'ignorance, tout comme la structure de base ainsi que les institutions qui en découlent, se doivent d'être neutres par rapport à la promotion des conceptions compréhensives du bien. En ce sens, elles ne doivent pas avantager aucune conception de la vie bonne afin que les individus puissent y adhérer librement. Young souligne néanmoins que les intérêts privés ont généralement tendance à se concentrer en impérialisme culturel. Effectivement, l'inclinaison culturelle des individus renforce l'hégémonie normative des groupes dominants puisqu'aucun mécanisme institutionnel ne s'y oppose. Comme nous le constaterons au prochain chapitre, Young est d'avis que c'est le rôle des institutions d'éviter le renforcement de l'impérialisme culturel. En ce sens, la justice ne devrait pas être déterminée par des êtres utopiques dont l'impartialité est assurée par un voile d'ignorance, mais plutôt par des êtres réels dont la différence est inévitable et même souhaitable. Young soutient donc qu'il est justifiable que certaines politiques publiques aient pour objectif de réduire le pouvoir de l'impérialisme culturel en facilitant la valorisation des conceptions de la vie bonne marginalisées, et elle défend donc la non-neutralité de ces dernières. Sans cette capacité des institutions à agir sur le plan de la justice représentationnelle, le cadre neutraliste proposé par Rawls est condamné à permettre, et même à renforcer, l'impérialisme culturel étant donné que l'État est dans l'impossibilité institutionnelle de valoriser les représentations stigmatisées.

Young ira plus loin en soutenant que l'idéal d'impartialité découlant du cadre neutraliste contribue également au renforcement de l'impérialisme culturel. « The ideal of impartiality legitimates hierarchical decision-making and allows the standpoint of the privileged to appear as universal. The combination of these function often leads to concrete decisions that perpetuate the oppression and disadvantage of some groups and the privilege of others. »¹⁶⁸ Effectivement, il s'ensuit que la seule existence de la domination culturelle d'un groupe sur les autres est suffisante afin d'expliquer la puissance normative de l'impartialité, et donc d'affirmer que les positions prises sous l'égide de ce principe avantagent finalement l'impérialisme culturel. Inconsciemment et par la faute de cet idéal, les individus participeraient donc collectivement à la dissimulation de leurs différences au profit de l'unité de la majorité normative. C'est en ce sens que la posture critique de Young à l'égard du cadre neutraliste rawlsien prend tout son sens, elle affirme ainsi: « Because I understand critical theory as starting from a specific location in a specific society, I can claim in this writing to be neither impartial nor comprehensive. I claim to speak neither for everyone, to everyone, nor about everything. »¹⁶⁹ La *politique de la différence* soumet donc un argument supplémentaire pour notre positionnement théorique puisque, comme nous le verrons au prochain chapitre, elle rend possibles des solutions s'adaptant mieux aux enjeux soulevés par le déploiement de l'IA.

3.3 L'interférence identitaire

Comme nous l'avons expliqué au début du chapitre, une action, ou l'action d'un algorithme, ne peut être déclarée injuste qu'en fonction de l'application d'une norme de justice. Une même décision algorithmique peut effectivement être considérée comme juste et injuste dépendamment de la norme sélectionnée pour la définition de l'enjeu en question. En acceptant l'hypothèse du phénomène de la performativité algorithmique, nous nous sommes concentrés sur l'oppression de l'impérialisme culturel puisqu'il nous permet la théorisation des injustices représentationnelles. Si la norme de justice associée à l'oppression était celle du développement personnel chez Young, elle deviendra celle de la non-interférence identitaire pour cet exposé. En ce sens, les algorithmes seront dorénavant considérés injustes lorsqu'ils interféreront négativement avec la construction identitaire

¹⁶⁸ YOUNG, Iris Marion, *Op. cit.*, p. 116.

¹⁶⁹ *Ibid.*, p. 13.

des individus, c'est-à-dire lorsqu'ils généreront des contraintes pour le développement de l'identité. Cette notion d'interférence est importante puisqu'elle nous permet de détecter la force contraignante des algorithmes tout en justifiant, comme nous allons le faire au dernier chapitre, une certaine pression émancipatrice. Pour la fin de ce chapitre, nous démontrerons que la force performative émergeant de l'accumulation des décisions algorithmiques amplifie les injustices structurelles puisqu'elle interfère avec la construction identitaire. Pour ce faire, nous poserons la question de la neutralité des algorithmes, des biais algorithmiques ainsi que des effets du *clustering* sur l'identité.

3.3.1 Le mythe de la neutralité algorithmique

L'une des idées ayant grandement contribué à l'acceptation sociale du déploiement de l'IA est la supposée neutralité des décisions algorithmiques. Dans le cadre de la justice digitale par exemple, il est compréhensible de concevoir que le fonctionnement mathématique des algorithmes peut sembler faire preuve de neutralité. Effectivement, puisque ces derniers sont des processus décisionnels automatisés dépourvus d'intentionnalité, ils devraient forcément être neutres, et donc impartiaux, tandis que les juges appartenant à des réalités sociohistoriques distinctes sont dans l'impossibilité ontologique de l'être. Bien que ce raisonnement puisse paraître valable à première vue, nous emploierons la notion de neutralité de Rawls afin de réfuter cette proposition. Étant donné le parallèle qu'il est possible d'établir entre une décision politique et une décision algorithmique, nous évaluerons la neutralité des algorithmes à la mesure de la neutralité des conséquences, des buts, et des justifications.

Tout comme les décisions politiques sont dans l'impossibilité d'être neutres selon la neutralité des conséquences, il en va de même pour les décisions générées par les algorithmes. Elles auront inévitablement des répercussions dans le monde social, et pour cette raison, elles ne sont pas neutres. Pour ce qui est de la neutralité des buts, Rawls soutenait que les décisions politiques pourraient être considérées neutres si leurs objectifs étaient neutres. Néanmoins, il est impossible d'unifier une intentionnalité partagée par plusieurs acteurs politiques, et même au niveau individuel, l'accessibilité à l'intentionnalité est compromise par l'intériorité de la personne. Analogiquement, la programmation des algorithmes dépend habituellement de la collaboration de plusieurs programmeurs et l'enjeu de l'unicité de leur intentionnalité est donc encore une fois soulevé. De plus,

l'intentionnalité d'un programmeur solitaire est également intérieure et donc inaccessible. Et même s'il était hypothétiquement envisageable d'attribuer une intentionnalité propre aux algorithmes, cette intentionnalité artificielle serait inintelligible. Tout comme lors de l'examen rawlsien de la neutralité, la seule neutralité sur laquelle il est possible de se fier est donc, par élimination, celle de la justification, c'est-à-dire d'affirmer qu'une décision est neutre lorsque les justifications des buts de cette dernière le sont. Appliqué à la réalité algorithmique, cela impliquerait que les programmeurs devraient justifier les décisions générées par les processus décisionnels automatisés, mais encore une fois, l'inintelligibilité de ces derniers rend la tâche extrêmement complexe. De surcroît, même si les programmeurs configurent initialement les algorithmes, ces derniers se transforment en fonction des données qu'ils intègrent, et c'était pour cette raison que nous avons qualifié ces derniers d'*eventful*, ou d'accomplissement en devenir. En acceptant que les algorithmes soient dynamiques et qu'ils s'érigent en complexes algorithmiques, il devient impossible, même pour les programmeurs, de justifier leurs décisions algorithmiques. Et même s'il devenait possible d'accéder à l'intelligibilité de ces derniers grâce à des éventuels et hypothétiques procédés de cartographie algorithmique, les décisions générées demeureront variables, car elles seront continuellement modifiées par l'apprentissage-machine. Maintenant que nous avons prouvé l'impossible neutralité des algorithmes en mobilisant la conception rawlsienne de la neutralité, et que nous avons par le fait même justifié l'insuffisance théorique de ce paradigme, il est intéressant de constater qu'au sein de la littérature, la non-neutralité des algorithmes est habituellement démontrée grâce à l'existence de biais algorithmiques.

3.3.2 Les biais algorithmiques

À la base, un biais est une direction oblique s'écartant de la direction principale, et généralement, il est question de biais cognitifs, renvoyant aux « modes intuitifs de pensée qui déforment (biaisent) le raisonnement logique et induisent des croyances erronées »¹⁷⁰, ou de biais méthodologiques désignant le non-respect d'un protocole. Quant aux biais algorithmiques, ils représentent l'entièreté des déviations pouvant déformer les décisions générées par des algorithmes lors des différentes étapes de la programmation. Deux sources

¹⁷⁰ KAHNEMAN, Daniel, *Thinking, Fast and Slow*, New York, 2011.

potentielles sont donc à distinguer afin d'expliquer l'infiltration des biais algorithmiques. D'une part, il y a les données, ainsi que la sélection de ces dernières, qui peuvent être biaisées, tandis que d'autre part, le fonctionnement algorithmique lui-même peut également l'être. Afin d'expliquer cette réalité, nous nous appuyerons sur les travaux de Suresh et Guttag qui se positionnent précisément une tentative d'unification des différentes catégorisations de biais algorithmiques.¹⁷¹ Ces catégories se distinguent selon l'étape à laquelle chacun des biais est introduit au sein de l'algorithme. Considérant globalement le cycle de vie d'un algorithme en apprentissage machine, ce dernier est séparé en deux phases majeures. En premier lieu, la génération de données, incluant les étapes de la collection et de la préparation de donnée, représente le processus au cours duquel toutes les données sont rassemblées et préparées, grâce à des techniques d'échantillonnage, pour que les algorithmes puissent être entraînés et évalués.¹⁷² En deuxième lieu, la construction et mise en œuvre des modèles, incluant les étapes du développement, de l'évaluation, du post-traitement ainsi que du déploiement du modèle, font directement référence à la programmation algorithmique qui s'effectue grâce à une succession d'entraînements et d'évaluations menant ultimement aux résultats escomptés.¹⁷³ Ce résumé de l'apprentissage machine nous permettra de situer chronologiquement les six différentes catégories de biais algorithmiques théorisées par Suresh et Guttag.

Premièrement, le biais historique survient lors de l'étape de l'extraction de données puisqu'il existe en tant que construction sociohistorique normative décrivant le monde tel qu'il est. Objectivement, il est possible d'observer un taux de criminalité plus élevé au sein des communautés afro-américaines et un algorithme apprenant pourrait facilement inférer, en fonction de la compilation de données soutenant cette réalité, que cette population est plus violente que les autres. Il s'agirait donc d'un cas de discrimination découlant d'un biais historique. Deuxièmement, le biais de la représentation s'explique par les techniques d'échantillonnage appliquées aux données. Lorsque des populations sont sous-représentées ou sur-représentées par rapport à certaines caractéristiques réelles, les algorithmes s'entraînant avec ces données seront possiblement soumis au biais de la représentation.

¹⁷¹ SURESH et GUTTAG, *A framework for understanding unintended consequences of Machine learning*, 2020, p.2

¹⁷² *Ibid.*

¹⁷³ *Ibid.*

C'est notamment le cas de la reconnaissance faciale dont le degré de précision dépend de la représentation des diverses populations dans les banques de données servant à l'entraînement des algorithmes. Effectivement, lorsqu'un algorithme de reconnaissance faciale s'entraîne avec une majorité de visages blancs, il est moins précis pour reconnaître des visages noirs par exemple. À titre d'exemple, Noble relève justement une situation au sein de laquelle un algorithme de reconnaissance faciale de Google avait identifié des visages afro-américains à des singes, résultat certainement rendu possible à cause de la sous-représentation des minorités visibles dans les banques de données. Il s'avère en effet, que le biais de la représentation joue un rôle particulièrement important dans la reconduction, et le renforcement, de l'impérialisme culturel puisque les programmeurs ont généralement tendance à surreprésenter les groupes dominants et privilégiés desquels ils font partie¹⁷⁴. Troisièmement, le biais de la mesure fait référence à l'étape durant laquelle les étiquettes, ainsi que les particularités des données, sont sélectionnées. En reprenant le cas de figure de l'algorithme COMPAS, une analyse ultérieure a démontré que chaque arrestation subie par un individu était automatiquement étiquetée comme la perpétration d'un crime. Il s'agit ici d'une erreur de mesure puisqu'il est faux d'affirmer que toute arrestation implique automatiquement un crime, une proportion importante d'arrestations sont effectivement des erreurs. Cette faute d'étiquetage biaise donc l'interprétation des données en associant algorithmiquement un nombre d'arrestations élevé à un taux de récidive élevé, et donc à une peine allongée. Quatrièmement, le biais de l'agrégation survient lorsque des populations sont combinées de manières inappropriées, pour être utilisées lors de la construction du modèle. En fonction de la tâche algorithmique, l'articulation des sous-groupes constituant une population combinée doit être cohérente avec les particularités de chacun de ces groupes. Par exemple, il a été démontré que les différents groupes sociaux n'étaient pas affectés par le diabète dans des proportions similaires, et l'agrégation de différentes populations, pour en faire une base de données unifiée, avantageait finalement le diagnostic de certains groupes lors du traitement algorithmique de ces derniers. Cinquièmement, le biais de l'évaluation résulte de l'évaluation même du modèle qui s'effectue habituellement grâce à des indicateurs de

¹⁷⁴ HICKS, Marie, *Programmed Inequality, How Britain discarded women technologist and lost its edge in computing*, 2017, p. 3.

performance. Le choix de ces indicateurs et référents peut assurément biaiser les algorithmes étant donné qu'ils détermineront si l'évaluation du modèle en question est jugée satisfaisante. Certains indicateurs avantageront certains groupes au détriment des autres. Sixièmement, le biais du déploiement se produit lorsque les résultats algorithmiques d'un modèle sont interprétés et mobilisés incorrectement. Encore une fois, l'algorithme COMPAS, qui avait strictement été programmé pour calculer le taux de récidive des prisonniers, avait finalement été utilisé pour le calcul de la peine. La nature de l'exploitation d'un algorithme peut biaiser ses résultats.

Suresh et Gutttag ne sont pas les seuls à avoir catégorisé les biais algorithmiques et il est intéressant de constater que Mehrabi et al. en dénombre plus de vingt et trois types en 2019¹⁷⁵. Dans les limites de ce mémoire, nous nous limiterons aux interrelations entre les six catégories de biais mentionnés étant donné qu'elles sont suffisantes pour rejeter l'idéal de la neutralité algorithmique. Effectivement, les biais algorithmiques ont tendance à s'autorenforcer grâce au phénomène de la boucle de rétroaction (feedback loop), « which is a situation in which the trained machine learning model makes decisions that produce outcomes, and these very outcomes affect future data that will be collected for subsequent training rounds. [...] This feedback loop is not only existant between the data and the algorithm, but also some work analyzed the existence of this loop between the algorithms and the user interaction. »¹⁷⁶ Dans *The Rise of Big Data Policing : Surveillance, Race, and the Future of Law Enforcement*, Andrew Ferguson révèle que des complexes algorithmiques sont utilisés afin de prédire les endroits où les crimes ont plus de chances d'avoir lieu, ces endroits qualifiés de *red boxes*¹⁷⁷. Entraînés grâce à une multitude de données démographiques et judiciaires, ces processus décisionnels automatisés recommandent donc aux policiers d'intervenir dans des localisations précises afin d'effectuer des arrestations. Outre les enjeux éthiques qu'une telle situation soulève, il est intéressant de constater qu'à chaque fois qu'un policier arrête un individu grâce à la recommandation de l'algorithme en question, le modèle se renforce grâce à l'effet de la boucle de rétroaction. Effectivement et comme nous l'avons vu au premier chapitre,

¹⁷⁵ MEHRABI, Ninareh, et AL, *A survey on Bias and Fairness in Machine Learning*, 2019.

¹⁷⁶ MEHRABI, Ninareh, et AL, *Op. cit.*, p. 7.

¹⁷⁷ FERGUSON, Andrew, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, 2017, p. 70.

lorsqu'un algorithme atteint son objectif, lorsque sa prédiction est juste, il est récompensé. Au fur et à mesure, les prédictions convergeront avec celles qui sont vraies puisque l'algorithme est programmé selon un modèle d'optimalité, son objectif étant de générer des prédictions qui s'avéreront. Cet exemple précis explique l'articulation du biais historique et du biais de la représentation puisque d'une part, les données ayant permis la délimitation des zones à *risque* sont initialement des informations démographiques et judiciaires décrivant une réalité sociohistorique, et d'autre part, les populations afro-américaines sont surreprésentées dans les diverses données relatives à la prédiction des zones rouges. Ces biais algorithmiques auront tendance à se renforcer à chaque fois que les policiers effectueront des arrestations dans les zones rouges étant donné que la collection des nouvelles données confirmera collatéralement la surreprésentation démographique de certains groupes sociaux dans les zones dangereuses. Concrètement, le phénomène de la boucle de rétroaction s'illustre lorsque la réalisation des prédictions algorithmiques renforce les biais algorithmiques y ayant contribué.

Sans conteste, l'existence de biais algorithmiques, exacerbée par le renforcement rétroactif de ces derniers, confirme une fois de plus la non-neutralité des algorithmes. Suivant ce raisonnement, il est possible d'affirmer que l'autorenforcement des biais algorithmiques de représentation contribue à la consolidation de l'impérialisme culturel. Effectivement, une surreprésentation des groupes dominants a pour effet de normaliser l'univers symbolique leur étant associé, tandis qu'une sous-représentation des autres groupes marginalise et stigmatise leur réalité. Comme nous le verrons au prochain point, cette participation des complexes algorithmiques à l'édification de l'impérialisme culturel est injuste puisqu'elle génère une interférence identitaire.

3.3.3 Les boucles de rétroaction

Afin de soutenir que les complexes algorithmiques participe au renforcement de l'impérialisme culturel comme nous le supposons, il est nécessaire que la distinction entre les groupes sociaux et les différentes catégorisations algorithmiques s'effacent afin que les mondes social et virtuel convergent. Comme nous l'avons supposé au deuxième chapitre, c'est notamment en raison de la pression des réseaux sociaux sur la socialisation que cette convergence aura tendance à se renforcer. Sans cette interaction performative entre ces deux mondes, la performativité algorithmique serait impossible.

La majorité des problèmes auxquels sont soumis les complexes algorithmiques peuvent généralement être différenciés en deux catégories, les problèmes de classification ainsi que les problèmes de régression. La classification algorithmique implique une réponse qualitative, et donc catégorique, obtenue grâce à la classification des données. La classification d'images de chat et de chien, par exemple, coïncide avec les réponses catégoriques de l'algorithme. Quant à elle, la régression algorithmique implique l'attribution d'une valeur numérique à un problème, qu'il s'agisse de la détermination d'un prix, d'un taux de risque ou encore d'un ordre de priorité. Bien que l'objectif premier d'un algorithme de régression ne soit pas la classification algorithmique, il arrive assez fréquemment que la valeur numérique générée reflète finalement une classification. Afin d'expliquer ce phénomène, nous pouvons constater que les réponses de l'algorithme COMPAS génèrent une classification sociale tout en étant numériques. Si le taux de récidive est influencé par des biais algorithmiques historiques et de représentation, il est clair que les valeurs numériques générées les auront intégrés lors du calcul algorithmique. De plus, lorsque l'on considère les algorithmes en apprentissage profond, il devient impossible de distinguer clairement la frontière entre les tâches de classification et celles de régression puisqu'un algorithme à plusieurs couches détient la faculté de mobiliser plusieurs moyens pour résoudre une tâche et son inintelligibilité rendra cette distinction impossible. Une régression peut alors être basée sur une classification tout comme une classification peut être influencée par une régression. Nous croyons donc qu'il devient en ce sens possible de parler généralement de classification algorithmique sans que nous devions pour autant exclure totalement les autres tâches effectuées par les algorithmes, comme celle de la régression, qui contribue au renforcement de l'impérialisme culturel.

Cette contribution s'explique, selon nous, de plusieurs façons. À partir du moment où nous acceptons que les complexes algorithmiques classifient les populations, et qu'ils génèrent des *clusters*, alors le traitement biaisé de ces différents groupes sociaux avantage et désavantage forcément les groupes entre eux. De plus, il faut considérer que les êtres humains, qui accomplissent le travail de programmation et d'étiquetage, participent eux-mêmes à l'impérialisme culturel puisque les groupes dominants sont statistiquement

surreprésentés dans le domaine de l'IA.¹⁷⁸ Outre la sélection des données et le travail d'étiquetage, les données en elles-mêmes sont également porteuses de biais y participant. Maintenant que nous avons démontré qu'il existe vraisemblablement plusieurs canaux entre le déploiement de l'IA et le renforcement de l'impérialisme culturel, nous pouvons revenir à la convergence entre les groupes sociaux et les classifications algorithmiques.

3.3.4 La stigmatisation et la conformité

Au début du chapitre, nous avons présumé deux tendances majeures aux répercussions du phénomène de la performativité algorithmique, celle de la conformité ainsi que celle de la stigmatisation. Ces dernières avaient pour objectif de démontrer comment la construction performative de l'identité pouvait engendrer un accroissement des injustices structurelles. En acceptant que certains groupes d'individus bénéficient de l'impérialisme culturel, et que la représentation de leur culture et univers symboliques soient favorisés, alors nous devons accepter que d'autres groupes sont défavorisés selon ces termes, ce qui constitue une injustice représentation. Il s'ensuit que les groupes sociaux soumis à l'impérialisme culturel sont désavantagés en fonction de la représentation, et de l'acceptation, de leur univers symbolique. Selon nous, cette réalité sociale impliquerait deux tendances pour ces groupes opprimés, celle de la conformité ainsi que celle de la stigmatisation. Dans un premier temps, les individus appartenant aux groupes sociaux dont la représentation ne s'accorde pas avec l'idéal de l'impérialisme culturel pourraient, selon une logique d'assimilation, se dissocier des particularités symboliques de leurs groupes de façon à construire leur identité en accord avec la pression normative vécue. Cette proposition impliquerait collatéralement un renforcement de l'hégémonie culturelle des groupes dominants, ainsi qu'un effacement de la différence identitaire caractéristique des autres groupes. Dans un deuxième temps, il serait également possible, pour les individus représentant les groupes sociaux opprimés, de résister à cette pression normative en affirmant leurs différences. Ces derniers s'exposeraient toutefois à l'exercice stigmatisant de l'impérialisme culturel, c'est-à-dire qu'il leur deviendrait plus difficile de participer à la vie sociale et politique afin de valoriser leurs revendications. Avec le renforcement de cette tendance entretenu par le conformisme de certains, il deviendrait de plus en plus difficile

¹⁷⁸ HICKS, Marie, *Op. cit.*, p.3.

pour d'autres d'incarner une identité se différenciant de la conformité. Dans un cas comme dans l'autre, il est clair que nous pouvons affirmer que le phénomène de la performativité algorithmique, amalgamé à l'action de l'impérialisme culturel, constitue une interférence à la libre construction de l'identité puisque la stigmatisation, ainsi que la conformité, privilégient ultimement les identités s'accordant avec l'hégémonie culturelle des groupes sociaux favorisés. Avec les générations et sans la mobilisation de mesures préventives importantes, nous croyons qu'il est possible de prédire une uniformisation de la différence, laquelle impliquerait une exacerbation des injustices structurelles. Comme nous le constaterons au prochain chapitre, *la politique de la différence* de Young offre des pistes de solution à la problématique de l'oppression de l'impérialisme culturel, et nous croyons pouvoir exploiter le pouvoir performatif des algorithmes en ce sens.

4 Pour une construction identitaire sans interférence

S'il est correct d'affirmer que le phénomène de la performativité algorithmique participe à la construction identitaire des individus, cela n'implique pas pour autant qu'il soit, en lui-même, porteur d'injustices. Effectivement, c'est seulement à partir du moment où il contribue et consolide l'impérialisme culturel, et donc qu'il interfère avec la construction identitaire, que le pouvoir performatif des algorithmes doit être considéré comme injuste et donc néfaste. Selon l'hypothèse que nous démontrerons, nous croyons qu'il est possible d'exploiter l'effet de la performativité algorithmique afin de réduire l'impérialisme culturel. Si cette manipulation était possible, cela signifierait que la force performative des algorithmiques pourrait être utilisée afin de faciliter, d'une part, le respect de la norme de la non-interférence identitaire, et d'autre part, l'émancipation des individus marginalisés. Le phénomène de la performativité algorithmique dans un contexte d'impérialisme culturel constitue à la fois un problème, car il exacerbe les injustices structurelles, et du même coup une solution à ce problème puisqu'il est potentiellement exploitable. Effectivement, si nous étions capables d'orienter les biais algorithmiques afin qu'ils valorisent les univers symboliques marginalisés, la performativité algorithmique serait exploitée à l'avantage de la justice sociale.

Afin d'opérationnaliser cette démarche, nous commencerons par décrire la matérialisation des effets injustes de la performativité algorithmique en illustrant certains cas de figure. Après cela, nous analyserons la proposition de Young pour traiter de l'oppression de l'impérialisme culturel, laquelle soutient que l'affirmation de la différence, articulée selon les principes de la représentativité et de l'inclusivité, est émancipatrice pour les groupes opprimés. Nous aborderons ensuite l'enjeu de la représentation et de l'inclusivité dans le domaine de la programmation algorithmique. Il s'agit de démontrer que ces deux éléments sont à la source du renforcement algorithmique de l'impérialisme culturel, et qu'il s'agit parallèlement d'un champ d'action potentiel pour le diminuer. Étant donné l'impossible neutralité des algorithmes, nous proposerons finalement l'articulation de l'exploitation des processus décisionnels automatisés afin qu'ils œuvrent au contrebalancement des pressions normatives de l'impérialisme culturel. Nous concluons donc que la théorisation de la performativité algorithmique nous permet de manipuler les biais algorithmiques afin qu'ils participent à l'atteinte d'une plus grande justice sociale.

4.1 L'échelle d'analyse

L'objectif sous-jacent de ce mémoire est d'offrir une échelle d'analyse permettant d'expliquer globalement les répercussions du déploiement de l'IA sur la justice sociale. En explorant les travaux de Noble et de plusieurs autres d'ailleurs, il est possible de constater qu'une vaste majorité des analyses ayant pour objets les injustices algorithmes se limitaient à la description de ces dernières. Effectivement et pour reprendre le cas de figure de la *black girl* de Noble, ce dernier se concentre sur des situations précises au sein desquelles l'algorithme PageRank génère des résultats de recherche sexistes et racistes. Bien qu'une description approfondie de ces situations d'injustice soit nécessaire, elle ne permet pas d'en avoir une vue d'ensemble et de comprendre que ces cas individuels sont les symptômes systémiques d'oppressions plus grandes, notamment de celle de l'impérialisme culturel. Le même raisonnement s'applique aux cas suivants, soit ceux de la théorisation de la discrimination générée par les algorithmes de prédiction policière effectuée par Ferguson, de l'analyse des impacts injustes des technologies de gestion de la pauvreté sur les groupes économiquement désavantagés¹⁷⁹ de Virginia Eubanks, ainsi que la démonstration de la discrimination systémique des algorithmes de profilage publicitaire¹⁸⁰ d'Amit Datta. Ce que nous tentons ici de démontrer, c'est que si toutes ces situations réelles représentent des entraves à la justice sociale, elles doivent actuellement toutes être analysées séparément puisqu'aucun cadre théorique ne permet de comprendre l'articulation du nombre incommensurable de décisions algorithmiques générées à chaque seconde. En l'absence d'un tel cadre, nous sommes effectivement condamnés à l'analyse des occurrences singulières des injustices algorithmiques, sans jamais pouvoir accéder au phénomène global. Nous soutenons en ce sens que l'approfondissement du concept de la performativité algorithmique rend possible l'esquisse d'un cadre théorique permettant d'avoir une vue d'ensemble des injustices structurelles provoquées par les décisions algorithmiques.

En expliquant les effets injustes de la performativité algorithmique reliés à l'impérialisme culturel, il devient possible de situer chacun de cas mentionnés selon leur

¹⁷⁹ EUBANKS, Virginia, *Automating Inequality, How High-Tech tools profile, police, and punish the poor*, St. Martin's Press, New-York, 2018, p. 15.

¹⁸⁰ DATTA, Amit, Michael Carl TSCHANTZ, et Anupam DATTA, « Automated Experiments on Ad Privacy Settings », *Proceedings on Privacy Enhancing Technologies*, 2015, p. 92.

participation concrète à cette injustice. Conformément à la perspective globalisante que nous proposons, nous affirmons que l’oppression stigmatisante de l’algorithme PageRank participe au renforcement de l’impérialisme culturel, tout comme c’est également le cas des algorithmes de prédiction policière, des algorithmes de gestion de la pauvreté, et des algorithmes de profilage et de ciblage publicitaire puisqu’ils contribuent tous à l’émergence d’une force performative qui renforce le pouvoir normatif des groupes dominants. De plus, nous affirmons que les effets de ces algorithmes sont injustes du point de vue de la justice sociale puisqu’ils constituent des freins à l’émancipation des individus. La force de notre approche résiderait donc dans l’intégration d’un grand nombre d’enjeux soulevés par les algorithmes qui s’étendent à toutes les situations où des décisions algorithmiques catégorisent les individus. En ce sens, à chaque fois que des mécanismes de *clustering* automatisés catégorisent les populations, ces *clusters* sont susceptibles de participer performativement au renforcement de l’impérialisme culturel, lequel se traduit par une pression normative conformante, ou stigmatisante, sur la construction identitaire des individus, ce qui constitue une interférence identitaire. Étant donné la généralité de l’enjeu identitaire, et que le *clustering* est une pratique algorithmique très largement répandue, nous croyons proposer une solution plausible au problème de la fragmentation des enjeux de justice soulevés par les algorithmes.

4.1.1 Le dynamisme de l’IA

Une autre réalité avec laquelle l’analyse des injustices algorithmiques doit être conjuguée est celle du dynamisme intrinsèque de l’IA. Effectivement, c’est en 2018 que Noble publie son ouvrage théorisant l’oppression de l’algorithme de PageRank. Sa recherche sur les “black girls” et la pornographie remonte toutefois à 2012. Entre temps, il est tout à fait possible qu’une compagnie avec les moyens de Google corrige certains résultats de recherche problématiques afin de préserver une image publique positive. À ce jour, force est de constater que les associations problématiques du moteur de recherche de Google se font de plus en plus rares. Cette réalité corrobore donc la définition qualifiant les complexes algorithmiques d’*eventful* ou des accomplissements en devenir. En ce sens, les complexes se transforment constamment, ils existent selon plusieurs versions et permutations, et il devient pour ces raisons, extrêmement difficile pour les chercheurs de documenter, d’analyser ou encore de prévoir chacune des occurrences représentant des

injustices algorithmiques. Dans les faits, la modification du fonctionnement d'un algorithme est souvent plus rapide que la théorisation de son occurrence. Pour revenir à l'exemple de l'algorithme PageRank de Noble, il est certainement profitable pour Google de maintenir son statut de source d'information fiable et neutre, et il lui devient donc justifiable d'investir massivement dans le maintien de cette apparence. La modélisation continue des algorithmes déterminant l'ordre d'affichage peut être manipulée de façon à ce que les résultats ne reconduisent plus de biais idéologiques manifestement stéréotypés, mais comme nous l'avons démontré plus tôt, il est impossible d'atteindre la neutralité algorithmique. Puisque les biais algorithmiques sont inhérents aux décisions générées, l'objectif de Google ne peut être l'atteinte d'une neutralité algorithmique, mais constitue plutôt celle d'une normalité socialement acceptable. Cette position normative de Google participe au renforcement de l'impérialisme culturel étant donné sa prétention à la neutralité. Si les représentations injustes décrites dans les travaux de Noble ne sont plus générées actuellement, cela n'implique pas pour autant que l'algorithme PageRank est dorénavant neutre et juste. Ainsi, nous soutenons donc que le dynamisme de l'IA exige la conceptualisation d'une perspective globalisante pour que puisse être abordée la multiplicité des enjeux qui adviendront.

L'image du *Far Web* est très éloquent afin d'illustrer cette course mettant en compétition la vitesse à laquelle s'accumulent les innovations technologiques aux réglementations qui leur sont attribuées. Les recommandations publiées par *La Déclaration de Montréal* s'appuient sur des théorisations constamment décalées par rapport à l'évolution des phénomènes se matérialisant en IA. Ce déphasage est inévitable puisque l'accumulation des nouvelles applications algorithmiques entraîne forcément des chocs qui peuvent se répercuter sur la justice sociale, tandis que la compréhension approfondie de ces derniers nécessite du temps et des recherches. En contrepartie et à l'image des innovations technologiques, le nombre grandissant de travaux traitant de ces questions permet à la communauté scientifique d'être proactive face à l'anticipation des implications de l'IA sur nos vies. Les travaux effectués par Noble ont d'ailleurs propulsé la conceptualisation de la performativité algorithmique pour l'adapter au dynamisme des injustices algorithmiques. Avant d'y arriver, nous reviendrons sur certaines propositions de Young.

4.2 L'affirmation de la différence

Au dernier chapitre, nous avons présenté le cadre théorique de la *politique de la différence* de Young sans pour autant exposer les pistes de solution qu'elle propose. Effectivement, la posture critique de cette dernière suppose une solution pratique à une problématique appartenant à une réalité sociohistorique particulière. De manière récapitulative, nous nous sommes concentrés sur l'enjeu non distributif de la culture et de son affiliation naturelle avec l'oppression de l'impérialisme culturel. Nous jugeons que le déploiement de l'IA, s'articulant dans la multiplicité des décisions algorithmiques générées, est un enjeu appartenant à cette catégorie étant donné sa participation au renforcement du pouvoir normatif de l'impérialisme culturel. Comme nous l'avons soutenu, c'est le phénomène de la performativité algorithmique qui permet d'expliquer cette participation, et son exploitation qui nous permettra potentiellement d'inverser son effet négatif.

4.2.1 La différence de traitement

Si cette idée peut paraître difficilement réalisable à première vue, cette dernière sera clarifiée grâce à son analogie avec la discrimination. En effet, Young relativise cette notion en déclarant que la discrimination ne constitue pas une injustice en soi, mais qu'elle devient plutôt injuste lorsqu'elle contribue à l'oppression. « [Cases of discrimination] have been wrong, that is, because they have contributed to and helped enforce oppression. If discrimination serves the purpose of undermining the oppression of a group, it may be not only permitted, but morally required. »¹⁸¹ L'exploitation de la performativité algorithmique, tout comme celle de la discrimination, implique la distinction entre l'objet et la nature de son utilisation. Dans son analyse, Young affirme que la discrimination est juste lorsqu'elle freine l'oppression. Si elle peut l'affirmer, c'est parce que l'objet de la justice est l'oppression, et que la discrimination ne constitue pas en elle-même une oppression puisqu'elle peut y participer positivement ou négativement. La discrimination est donc injuste lorsqu'elle contribue minimalement à l'une des cinq facettes de l'oppression. C'est le cas lorsqu'elle renforce par exemple l'impérialisme culturel en défavorisant les groupes minoritaires face aux groupes dominants, tandis qu'elle est juste lorsqu'elle affaiblit l'impérialisme culturel en favorisant les groupes y étant initialement

¹⁸¹ *Ibid.*, p.197.

soumis. Nous parlerons alors de discrimination positive, et le même raisonnement devrait selon nous s'appliquer à la performativité algorithmique.

4.2.2 Le rejet de l'impartialité

Selon le cadre théorique de Young, la justice sociale est représentée par la dissolution de l'oppression et de la domination. Elle ajoute que « democracy is both an element and condition of social justice. »¹⁸² La démocratie à laquelle fait référence Young, la démocratie en tant que *politique de la différence*, est un régime au sein duquel l'affirmation, et l'existence de la différence, sont nécessaires dans l'atteinte de la justice sociale. Pour bien expliquer cette idée, il faut revenir sur le rapport entre les idéaux de neutralité et d'impartialité, et le renforcement de l'impérialisme culturel. Comme nous l'avons déjà mentionné, Young s'oppose au *libéralisme politique* de Rawls dans le cadre d'une démocratie constitutionnelle puisque la doctrine libérale promeut une neutralité politique qui participe à la valorisation de l'impartialité, et par ce biais au renforcement de l'impérialisme culturel. Le pouvoir normatif de l'impérialisme culturel est effectivement une menace pour l'affirmation de la différence puisqu'il participe à sa dévalorisation en prescrivant une conformité normative. Si l'impérialisme culturel constitue une injustice de représentation, c'est parce qu'il participe activement à l'effacement de la différence en interférant avec sa représentation. Afin d'éviter cette tendance naturelle des démocraties libérales à renforcer l'impérialisme culturel, elle propose de fonder la démocratie sur l'affirmation, et même la célébration de la différence. Là où l'impartialité encourage l'effacement des différences individuelles au nom d'un idéal de neutralité inatteignable, l'affirmation de la différence supporte plutôt la normalisation de l'expression de la singularité au sein de la participation politique. Comme elle l'indique, « instead of a fictional contract [impartiality], we require real participatory structures in which actual people, with their geographical, ethnic, gender, and occupational differences, assert their perspectives on social issues within institutions that encourage the representation of their distinct voices. »¹⁸³ Si les structures participatives actuelles contribuent à l'impérialisme culturel en favorisant une représentation impartiale des individus, les structures participatives proposées par Young ont inversement pour objectif de limiter son pouvoir

¹⁸² YOUNG, Iris Marion, *Justice dans the politics of difference*, 1990, p. 91.

¹⁸³ *Ibid.*, p. 116.

normatif en sollicitant la représentation de la différence des groupes y étant soumis, laquelle constituerait un contre-pouvoir normatif à l'impérialisme culturel. De manière concrète, Young affirme donc que le pouvoir normatif des groupes dominants doit être contrebalancé par l'affirmation de la différence des groupes minoritaires. Néanmoins, afin que les groupes dominés puissent être en mesure d'affirmer leurs différences dans les processus de participation et de prise de décision démocratiques, la poursuite de l'idéal d'impartialité doit être abandonnée.

4.2.3 La redéfinition de l'égalité

L'affranchissement de cet idéal, accompli par une révolution subjective et culturelle esquissée Young, permettrait d'affaiblir de l'impérialisme culturel.¹⁸⁴ Cette révolution est subjective, puisqu'elle vise la déconstruction des aversions inconscientes à la différence chez les individus, et elle est culturelle, puisqu'elle implique un idéal de participation inclusif de tous les groupes sociaux¹⁸⁵. Ces aversions empêchent la normalisation de la différence en perpétuant l'impérialisme culturel. « Unconscious racism, sexism, homophobia, ageism, and ableism occur not only in bodily reactions and feeling and their expression in behavior, but also in judgments about people or policies. »¹⁸⁶ Young est d'avis que le champ d'application de la justice doit s'étendre à toutes les formes d'oppression, considérant qu'elles peuvent provenir des actions, comportements et jugements conscients, mais que leur source peut également être inconsciente. S'il est effectivement questionnable d'interdire et de punir ce qui n'est pas intentionnel, Young soutient néanmoins que la prise de conscience des effets oppressants des réactions et habitudes inconscientes est nécessaire au rétablissement de la justice sociale¹⁸⁷.

Nous soutenons que la performativité algorithmique devrait œuvrer à la représentation positive de la différence, ce qui contribuerait à la déconstruction des aversions subjectives à son égard. Avant d'en arriver à l'opérationnalisation de cette solution, nous devons redéfinir l'idéal d'égalité puisque cette redéfinition sera nécessaire

¹⁸⁴ *Ibid.*, p. 124.

¹⁸⁵ *Ibid.*, p. 134.

¹⁸⁶ *Ibid.*, p. 134.

¹⁸⁷ *Ibid.*, p. 153.

à la justification de notre proposition, laquelle s'inscrit dans le prolongement de la révolution culturelle chez Young.

«The assimilationist ideal assumes that equal social status for all persons requires treating everyone according to the same principles, rules, and standards. A politics of difference argues, on the other hand, that equality as the participation and inclusion of all groups sometimes requires different treatment for oppressed or disadvantaged groups. To promote social justice, I argue, social policy should sometimes accord special treatments to groups. »¹⁸⁸

Young distingue effectivement l'égalité de traitement et l'égalité de participation et d'inclusion dont l'instauration est nécessaire pour atteindre de la justice sociale. Dépendamment des situations, l'application des mêmes principes, règles et standards peut contribuer au renforcement de l'impérialisme culturel en avantageant systématiquement certains groupes favorisés. Afin de l'illustrer, Young conteste l'idéal de la méritocratie¹⁸⁹ en soutenant qu'un traitement égal engendre finalement des inégalités se répercutant sur la justice sociale. En contrepartie, la différence de traitement est justifiée puisqu'elle permet l'atteinte de l'égalité de participation et d'inclusion des individus dans la sphère politique et sociale, laquelle rend possible la surreprésentation des intérêts de la diversité, et ultimement, la limitation de l'impérialisme culturel¹⁹⁰ ? La discrimination positive s'inscrit donc dans l'atteinte d'un objectif d'égalité de participation et d'inclusion, et est donc jugée positivement du point de vue de la politique de la différence même si elle contrevient à l'égalité de traitement. Suivant ce raisonnement, nous argumenterons que l'exploitation de la performativité algorithmique est justifiable lorsque l'égalité de participation et d'inclusion est favorisée.

4.2.4 Les politiques fondées sur la reconnaissance des groupes¹⁹¹

Chez Young, la discrimination positive appartient à la catégorie des politiques fondées sur la reconnaissance des groupes, lesquelles sont nécessaires à l'affirmation de ces derniers¹⁹². Au niveau politique, l'égalité de participation et d'inclusion est concrétisée démocratiquement dans la notion d'un contre-pouvoir culturel, soit celui de la coalition

¹⁸⁸ *Ibid.*, p. 158.

¹⁸⁹ *Ibid.*, p. 200.

¹⁹⁰ *Ibid.*, p. 158.

¹⁹¹ Il s'agit d'une traduction de l'anglais de la notion de *group-conscious policies* théorisée par Iris Marion Young dans *Justice and the politics of difference* p. 173-176.

¹⁹² *Ibid.*, p. 174.

arc-en-ciel¹⁹³. L'instauration de cette institution politique, dont l'objectif est de rétablir une justice représentationnelle dans l'arène politique, reconnaît le rapport de force asymétrique existant entre l'impérialisme culturel et les intérêts atomisés des différents groupes sociaux. C'est le dépassement de cette opposition entre l'unicité et la multiplicité, entre la norme et la différence, que tente précisément de concrétiser ce type de coalition. À l'antipode de la coalition traditionnelle dont les bases sont fondées sur l'unification des intérêts impartiaux, et collatéralement sur la dissolution des intérêts issus de la diversité, la coalition arc-en-ciel représente le véhicule au sein duquel peuvent coexister les intérêts multiples des différents groupes la constituant¹⁹⁴. Dans ce type de coalition, la coexistence de la différence devient possible grâce à la similarité des expériences d'oppressions vécues par les groupes sociaux. Bien que ces situations soient singulières, et que l'émancipation nécessite des ajustements aux besoins particuliers, il est tout de même possible de discerner une convergence émancipatrice propre aux coalitions arc-en-ciel. De plus, le potentiel démographique important de ces associations, rendu possible grâce à leur nature inclusive, fait en sorte qu'elles constituent un contre-pouvoir représentationnel vraisemblablement suffisant pour s'opposer à l'impérialisme culturel. Théoriquement, tous les groupes soumis à l'impérialisme culturel pourraient participer à l'affaiblissement de son hégémonie normative et représentationnelle. En évitant le cannibalisme entre les représentations des groupes sociaux y étant soumis grâce à la coalition arc-en-ciel, il devient possible de regrouper leur force représentationnelle afin qu'elle participe à la remise en question de l'impérialisme culturel et de son pouvoir normatif, ce qui permettrait de tendre vers l'idéal de la justice sociale chez Young.

4.3 L'inclusivité et la représentativité

Le principe de l'égalité de représentation et d'inclusion doit donc encadrer la participation démocratique afin que l'oppression puisse être limitée. C'est en ce sens que l'édification d'une coalition arc-en-ciel constituerait un contre-pouvoir normatif à l'impérialisme culturel. Si l'enjeu du déploiement de l'IA diffère des considérations politiques initialement soulevées par Young, nous démontrerons que l'égalité de représentation et d'inclusivité représente un idéal vers lequel devrait tendre l'exploitation

¹⁹³ *Ibid.*, p. 188-189.

¹⁹⁴ *Ibid.*, p. 188-189.

de la performativité algorithmique afin que la norme de la non-interférence identitaire puisse être préservée. Nous soutenons qu'il est possible d'établir un parallèle entre l'atteinte de l'égalité de représentation et d'inclusion dans la participation démocratique et son atteinte au sein du déploiement de l'IA. Il serait possible d'exercer une influence dirigée sur les biais algorithmiques en instaurant une politique de représentation inclusive au sein de la programmation. Nous expliquerons donc comment notre conception de la représentation inclusive s'articule avec l'exploitation du phénomène de la performativité algorithmique.

4.3.1 Les enjeux de représentativité au sein de l'IA

Pour commencer, il est impératif de prendre conscience de la faible diversité du milieu de l'IA. Dans *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*, Marie Hick explique comment la division genrée du travail a engendré et naturalisé une représentation quasi exclusivement masculine dans le développement de l'ordinateur. Cette réalité explique socialement et historiquement l'inégalité représentationnelle dans l'univers de l'IA actuelle¹⁹⁵. Effectivement, « rappelons que les femmes représentent près de la moitié de l'humanité, mais à peine 33% des personnes du secteur numérique (12% seulement, si l'on écarte les fonctions transversales et supports). »¹⁹⁶ L'observatoire international sur les impacts sociétaux de l'IA et du numérique soulève que « la représentation des femmes ainsi que des groupes sous-représentés est très faible »¹⁹⁷ dans le secteur numérique et de l'IA. Comme démontré précédemment, les biais algorithmiques participant à l'oppression de l'impérialisme culturel s'expliquent notamment grâce à la reconduction des biais des concepteurs à travers l'étape de la programmation, ainsi qu'en fonction des méthodes d'échantillonnage préconisées lors de la sélection des données. En ce sens, nous soutenons que le travail humain au fondement des algorithmes, en termes de programmation, de sélection et d'étiquetage des données, peut orienter les biais algorithmiques afin qu'ils participent à la limitation de l'impérialisme culturel. En contrepartie, il est important que

¹⁹⁵ HICKS, Marie, *Programmed Inequality, How Britain discarded women technologist and lost its edge in computing*, 2017, p. 3.

¹⁹⁶ VILLANI, Cédric, *Donner un sens à l'intelligence artificielle, Pour une stratégie nationale et européenne*, 2018, p. 163.

¹⁹⁷ OBVIA, *Équité, diversité et inclusion*, En ligne au : <https://observatoire-ia.ulaval.ca/edi/>, consulté le 25 mai 2022.

ce travail humain soit accompli par des représentants de la diversité afin que les biais reconduits s'opposent au pouvoir normatif de l'impérialisme culturel. Si les algorithmes participent aujourd'hui au renforcement de l'impérialisme culturel, c'est premièrement parce que le secteur numérique de l'IA est démographiquement dominé par les groupes sociaux le représentant, et deuxièmement parce que l'impartialité exigée de tous contribue collatéralement à la réification de son pouvoir. Pour ces raisons, la réduction de l'impérialisme culturel nécessite d'une part, l'instauration d'une politique de représentation inclusive des différents groupes sociaux dans le secteur de l'IA, et d'autre part l'édification d'une structure permettant l'affirmation de leurs intérêts. C'est donc grâce à cette double intervention au sein du développement de l'IA qu'il deviendrait possible d'utiliser l'effet de la performativité algorithmique afin qu'elle contribue à l'atteinte d'une plus grande justice sociale.

4.3.2 L'éducation et la sensibilisation

Dans son ouvrage ayant pour ambition, comme son titre l'indique, de *Donner un sens à l'intelligence artificielle*, Cédric Villani soutient effectivement qu'une « éducation du plus grand nombre aux principes de l'IA et des algorithmes est primordiale »¹⁹⁸ dans l'instauration d'une politique d'inclusion dans ce domaine. Nous sommes évidemment d'avis qu'une plus grande connaissance de l'IA, dont les fondements sont encore incompréhensibles pour la majorité, démocratisera son accessibilité pour tous. Puisque l'objectif est premièrement d'ériger une politique de représentation inclusive dans le développement de l'IA, nous soutenons que la première étape afin d'y parvenir devrait effectivement être de permettre son attractivité grâce à l'éducation. Même s'il est vrai que la thématique de l'IA est un sujet dont l'importance s'est récemment popularisée au sein de la communauté universitaire, il n'en demeure pas moins que les frontières de cette tendance contribuent à l'homogénéité de ses adeptes. Pour une plus grande diversité, il est nécessaire que tous les individus de toutes les provenances socioéconomiques et culturelles aient accès à un enseignement, ne serait-ce que rudimentaire et dépassant donc les limites universitaires, des bases du fonctionnement de l'IA et des enjeux qu'il soulève. Nous limitant à la structure académique québécoise pour les besoins de nos propos, et soutenant

¹⁹⁸ *Ibid.*, VILLANI, Cédric, p. 163.

que l'application de cette politique devrait être globalisée, nous jugeons que l'éthique de l'IA devrait s'ajouter à la liste des matières obligatoires à enseigner au niveau secondaire afin que tous puissent y être sensibilisés. Plus grande sera la conscientisation des impacts sociaux de l'IA, et plus diversifiée et inclusive sera la représentation des générations futures de chercheurs s'y intéressant. Si cette piste de solution s'inscrit dans une temporalité longue, c'est parce que nous soutenons que le déploiement de l'intelligence est encore à un stade embryonnaire. La robotique, la réalité virtuelle et l'automatisation sont tous des domaines dont le potentiel continuera certainement à s'étendre. Les nouvelles opérationnalisations des algorithmes impliquent assurément de nouvelles répercussions sociales, et c'est pour cette raison que nous recommandons une intégration de ces considérations au sein du parcours académique obligatoire.

4.3.3 La diversité et la discrimination positive

Bien que l'intérêt des individus de tous les milieux socioéconomiques et culturels aux enjeux de l'IA soit une condition nécessaire à une représentation inclusive dans ce domaine, elle n'est pas suffisante afin d'y parvenir. Il devient nécessaire de mettre en œuvre une politique de discrimination positive pour que les membres des groupes sous-représentés obtiennent de meilleures chances de participer au développement de l'IA, ce qui engendrerait une plus grande diversité. De manière concrète, il pourrait être exigé, grâce à des quotas gouvernementaux, que les entreprises responsables de la programmation algorithmique engagent des individus dont l'appartenance est mixte et diversifiée. Les politiques de ce type existent déjà dans plusieurs milieux et leur opérationnalisation peut assurément s'inspirer de ce qui existe à ce jour. De toute évidence, certaines compétences sont assurément requises pour toutes les professions, et il ne s'agit pas de faire abstraction de ces dernières en priorisant la diversité au coût de l'expertise. Il s'agit plutôt, et notamment grâce aux mesures qui doivent être mobilisées en éducation, d'élargir la diversité du bassin d'individus détenant cette expertise afin qu'il devienne possible de favoriser la compétence issue de la diversité. De manière générale, les exigences en termes de représentation des différentes minorités visibles au sein d'entreprises sont déterminées institutionnellement. Même s'il est souhaitable que des paramètres soient définis à cette fin, cette question ne sera pas traitée dans ce mémoire. À ce stade, notre proposition est générale et affirme plutôt qu'une plus grande mixité dans le domaine de l'IA, rendu

possible grâce à une éducation en la matière, accessible à tous, constitue une avancée dans la lutte contre les biais algorithmiques participants à l'impérialisme culturel.

4.4 La Coalition Arc-en-ciel Représentative de l'Intelligence Artificielle (CARIA)

Jusqu'à maintenant, ce que nous avons proposé est très largement répandu au sein de la communauté, et c'est d'ailleurs pour cette raison que nous n'avons pas jugé pertinent d'approfondir sur le sujet. En contrepartie, nous croyons qu'il serait possible d'augmenter démographiquement la diversité de manière à convertir leurs intérêts en biais algorithmiques, ce qui constituerait une utilisation de la performativité algorithmique limitant l'impérialisme culturel. Pour pouvoir opérationnaliser cette stratégie, nous utiliserons le concept de la coalition arc-en-ciel, que nous rebaptiserons Coalition Arc-en-ciel Représentative de l'Intelligence Artificielle (CARIA). À l'image des critères de représentativité et d'inclusivité que Young applique à l'espace démocratique, nous soutenons que le domaine de l'IA se doit d'être soumis aux mêmes critères afin que son développement respecte les exigences de la justice sociale. De la même manière, l'idéal d'impartialité doit y être abandonné afin de limiter le pouvoir normatif de l'impérialisme culturel. Pour cette raison, l'inclusion des différents groupes minoritaires dans le projet de l'IA est une condition nécessaire, mais non suffisante à notre solution. Elle est nécessaire parce que la constitution d'une coalition arc-en-ciel implique nécessairement l'intégration d'experts issus de la diversité, mais elle est insuffisante puisque la seule intégration de ces individus au sein du projet de l'IA ne garantit pas qu'ils seront en mesure de faire valoir la divergence de leurs intérêts. C'est pour répondre au défi de la prise en considération des intérêts des groupes dominés qu'intervient la CARIA. Son articulation devrait selon nous être la suivante. Pour commencer, nous croyons qu'il est impossible et irréaliste d'exiger que toutes les entreprises de l'IA intègrent individuellement une coalition de ce type au sein de leurs équipes de programmation. Nous soutenons en revanche qu'il soit envisageable de fonder une coalition arc-en-ciel publique et indépendante, dont le mandat est de constituer un contre-pouvoir normatif lors des processus de programmation et de sélection des données, et dont la mission est la mise en valeur des intérêts issus de la diversité. Si le portrait de cette organisation est évidemment encore flou à ce stade hypothétique, nous tenterons tout de même d'en définir les contours.

Selon notre vision du déploiement responsable de l'IA, nous croyons que le rôle de la CARIA devrait être d'orienter le développement, ainsi que la réglementation, des complexes algorithmiques sociaux performativement actifs. Puisqu'il est établi que certains processus décisionnels automatisés ont des répercussions sur la justice sociale, et que leur accumulation répétitive génère un effet performatif sur la construction identitaire, il est justifiable qu'ils soient soumis à l'autorité d'une instance supérieure avant d'être déployés. Cette autorité ne se limiterait pas à une approbation ou à une désapprobation des algorithmes, mais elle s'impliquerait également au cœur de leur programmation afin qu'ils répondent à certaines exigences que nous déterminerons. Les algorithmes devraient donc être réglementés selon certaines normes en fonction du champ d'action des décisions algorithmiques, de la quantité et de la répétition de ces dernières, ainsi qu'en fonction de ce qu'elles génèrent. Concrètement, le rôle de la CARIA se modulerait donc à la mesure de ces trois critères, et culminerait dans l'exploitation de la performativité algorithmique.

4.4.1 La classification des complexes algorithmiques

4.4.1.1 *Les algorithmes purs et les algorithmes sociaux*

Premièrement, les algorithmes devraient être classifiés selon le champ d'action dans lequel leurs décisions s'inscrivent. Étant donné la focalisation de la CARIA sur la justice sociale, seuls les complexes algorithmiques ayant des effets sociaux devraient être réglementés durant leur conception, leur développement, et leur déploiement, tandis que les autres devraient être libérés de telles exigences. Nous proposons donc une délimitation entre les algorithmes que nous qualifierons de purs et de sociaux. D'une part, les algorithmes purs représentent des outils dont la force de calcul est utilisée dans la résolution des problèmes appartenant à la catégorie des sciences fondamentales, c'est notamment le cas des calculs qu'accomplirait une calculatrice ou un algorithme de prévisions météorologiques. D'autre part, les algorithmes sociaux ont comme objet principal l'être humain et son comportement, ce qui implique que les décisions qu'ils génèrent ont des répercussions sur son existence sociale. Si cette distinction définitionnelle peut paraître simple a priori, elle est difficilement discernable dans la pratique puisque l'usage de certaines applications a priori pures peut générer des répercussions sociales. C'est notamment le cas du moteur de recherche de Google dont la focalisation n'est pas sur l'être humain au premier abord, mais dont les répercussions ont une résonance sociale. Ce

premier rôle de classification de la CARIA est d'une importance capitale étant donné qu'il constitue la première ligne de défense dans l'interprétation des dérives sociales que peuvent provoquer les processus décisionnels automatisés. Tous les complexes algorithmiques sociaux devraient donc préalablement obtenir le droit d'être déployés dans la société, et cette accréditation devrait être accordée par la CARIA. Considérant le fait que des algorithmes sociaux se sont déployés librement au cours des dernières années, il serait donc nécessaire de classer rétroactivement les complexes algorithmiques afin de déterminer ceux qui sont actifs socialement. Une fois la classification des algorithmes ayant une incidence sur la justice sociale accomplie, il devient ensuite nécessaire de mesurer cette dernière.

4.4.1.2 L'évaluation de la performativité

Deuxièmement et dans le prolongement de cette volonté de classification des algorithmes sociaux, la CARIA déterminerait si ces derniers participent au phénomène de la performativité algorithmique, que ses répercussions soient bénéfiques ou néfastes du point de vue de la construction identitaire. Pour ce faire, elle sera amenée à évaluer l'ampleur de la répétition des décisions qu'ils génèrent. Pour reprendre l'exemple du moteur de recherche de Google, l'architecture de son fonctionnement occasionne l'émergence d'un pouvoir performatif puisque la répétition de chaque recherche générée est porteuse d'un contenu représentationnel pouvant potentiellement interférer avec la construction identitaire. Il serait faux d'affirmer que tous les algorithmes sociaux sont performatifs, et il devient donc nécessaire à ce stade d'isoler ceux qui le sont puisqu'uniquement ceux-ci pourront être exploités dans la limitation de l'impérialisme culturel. Néanmoins et comme nous pouvons le constater auprès de la très grande majorité des complexes algorithmiques sociaux actuellement déployés, l'accumulation et la répétition des décisions algorithmiques constituent généralement la norme puisque c'est de cette façon que les algorithmes s'entraînent et se perfectionnent. Pour que le microciblage publicitaire soit par exemple effectif, les techniques de marketing exigent qu'il soit répété, et c'est ainsi qu'il devient performativement contraignant. Comme nous l'avons montré précédemment, la performativité algorithmique n'est pas néfaste en soi, car ses effets varieront en fonction des biais algorithmiques reconduits lors de l'accumulation des décisions générées, et c'est donc lors de la prochaine étape qu'ils devront être évalués.

4.4.1.3 *L'évaluation du contenu représentationnel*

Une fois que les algorithmes sociaux et performatifs auront été isolés, il sera ensuite nécessaire de déterminer la nature des effets générés sur la justice sociale. Bien que la focalisation de ce mémoire porte sur la construction identitaire, et que la norme de justice que nous privilégions soit celle de la non-interférence identitaire, il est possible d'imaginer une situation où la CARIA se concentrerait sur un autre enjeu soulevé par le phénomène de la performativité algorithmique. Pour notre propos, nous soutiendrons cependant que le pouvoir performatif des algorithmes devra être considéré comme injuste à partir du moment où il interfère avec la construction identitaire. L'interférence doit en ce sens être caractérisée par l'effet performatif contraignant que génère un algorithme sur la construction identitaire d'un individu. À cet égard, nous avons supposé que ces contraintes pouvaient se matérialiser par la stigmatisation et/ou la conformisation des contenus représentationnels soumis à l'impérialisme culturel. Dans les deux cas, la CARIA diagnostiquerait une situation d'interférence puisque la norme de la libre construction identitaire ne serait pas respectée. À ce titre, l'exemple de la *“black girls”*¹⁹⁹ de Noble représente évidemment une situation d'interférence puisque la construction identitaire des représentants de ces groupes subira l'effet performatif des algorithmes, lequel stigmatisera le contenu représentationnel leur étant associé. Le devoir de la CARIA sera donc d'évaluer les contenus représentationnels interférant avec la construction identitaire. C'est seulement une fois cette évaluation accomplie que pourra être entamée l'exploitation de la performativité algorithmique.

4.4.2 Les méthodes d'exploitation de la performativité algorithmique

4.4.2.1 *L'intégration des données*

Comme nous l'avons expliqué lors de la section sur les biais algorithmiques, leurs reconstructions survient à plusieurs étapes du déploiement des algorithmes. En ce qui concerne l'exploitation de la performativité algorithmique dans la réduction de l'impérialisme culturel, nous soutenons que la CARIA serait en mesure d'avoir une incidence directe sur trois étapes clés de la programmation afin de modifier l'orientation des biais algorithmiques. Dans un premier temps, la sélection des données en vue de l'entraînement d'un complexe algorithmique est une phase cruciale de son développement

¹⁹⁹ *Ibid.*

puisque'il s'agit du moment où sont intégrées les informations qui deviendront constitutives de l'apprentissage machine. Étant nécessairement biaisées, les données, ainsi que les techniques d'échantillonnage employées pour leur traitement, déterminent la nature des biais algorithmiques qui orienteront les processus décisionnels automatisés. S'il est par exemple possible d'expliquer le contenu représentationnel stigmatisant associé à la recherche "black girl" grâce à la surreprésentation des données l'associant à des résultats pornographiques, il est également possible de surreprésenter artificiellement son association à un contenu représentationnel valorisant en manipulant la nature, ainsi que la sélection, des données récoltées à cet effet. Il devient donc impératif que les individus jugeant de l'effet stigmatisant et valorisant d'un contenu représentationnel, soient des représentants du ou des groupes visés. Le rôle de ces représentants, membres de la CARIA, serait donc dans un premier temps d'identifier les contenus représentationnels stigmatisants, et dans un deuxième temps de déterminer la mise en valeur adéquate de ces derniers. Une fois cette tâche accomplie, la CARIA pourrait imposer un plan d'intégration des données aux complexes algorithmiques problématiques. Il consisterait en une réorientation des biais algorithmiques contribuant à la représentation stigmatisante de manière à ce qu'ils génèrent inversement un contenu représentationnel valorisant. Concrètement, cela veut dire que les membres de la CARIA, représentant pour notre exemple le groupe intersectionnel des "black girls", devraient premièrement déterminer si les résultats de recherches associés à la pornographie sont néfastes d'un point de vue représentationnel. Deuxièmement, ils devraient proposer la substitution d'un contenu représentationnel qu'ils jugent valorisant. On peut imaginer, par exemple, des résultats de recherche associés à des femmes noires présentant certaines valeurs dont la résonance est positivement perçue par la communauté. Troisièmement et pour rendre effectifs ces choix sur le marché de la justice représentationnelle, la CARIA pourrait forger artificiellement une base de données permettant l'intégration, ainsi que la reconduction, des biais désirés. Dans la pratique, plusieurs résultats de recherche associés à différents groupes sociaux peuvent être problématiques, et l'élaboration d'un plan d'intégration des données devrait corriger les multiples représentations négatives générées par le complexe algorithmique en question. La modélisation d'une base de données unifiée est donc nécessaire dans l'incorporation de données visant la mise en valeur des groupes visés.

4.4.2.2 L'étiquetage pour dé-biaisier

Dans un deuxième temps, selon les systèmes d'apprentissage machine utilisés lors de la programmation, les membres de la CARIA seraient amenés à participer au travail d'étiquetage des données afin qu'elles mettent en valeur les intérêts des représentants des groupes opprimés. Tout comme au point précédent, les intérêts véhiculés dans ce travail d'étiquetage dépendraient de la détermination des injustices représentationnelles puisque la tâche serait de constituer un contre-pouvoir normatif à l'impérialisme culturel. Constatant que l'étiquetage des données est généralement effectué par des individus issus de l'univers de la programmation, il est possible de supposer que le manque de représentativité de ces derniers participe au renforcement de l'impérialisme culturel. En ce sens, nous soutenons donc qu'il est impératif que les intérêts des représentants des groupes minoritaires soient surreprésentés au sein du travail d'étiquetage. Étant donné que les effectifs de la CARIA seront insuffisants pour constituer un contre-pouvoir démographique permettant le renversement de l'impérialisme culturel, il sera nécessaire de surreprésenter artificiellement le travail d'étiquetage de la CARIA de manière à ce que soient défendus les intérêts de la diversité.

4.4.2.3 Les critères d'évaluation

Dans un troisième temps, le champ d'action de la CARIA pourrait finalement s'étendre à la détermination des critères d'évaluation des complexes algorithmiques. Lorsque la performance d'un algorithme est évaluée uniquement selon des critères économiques, il va de soi que l'équipe de programmation n'accorde aucune importance à la reconduction des injustices représentationnelles. Nous soutenons donc que ce serait le rôle de la CARIA d'imposer l'intégration de certains critères permettant l'évaluation des algorithmes selon la nature des représentations qu'ils génèrent. Tout comme nous l'avons précédemment établi, les critères imposés à la programmation des algorithmes viseraient à valoriser les intérêts des représentants des groupes soumis à l'impérialisme culturel. Il pourrait par exemple être déterminé, selon un critère d'anti-sexualisation des individus, qu'aucun résultat de recherche ne doit être associé à des résultats pornographiques sans que l'utilisateur ne l'exige explicitement. Le respect d'un tel critère empêcherait effectivement les injustices relatives aux associations problématiques du moteur de recherche de Google. La mise en place de certains critères clés d'évaluation constituerait

une protection supplémentaire aux dérives potentielles des algorithmes en termes de justice représentationnelle, et elle permettrait donc le maintien de règles durables et objectives à cet effet.

Bien entendu, ces méthodes n'ont pas la prétention d'être exhaustives et invariables puisque la CARIA serait forcément amenée à se renouveler afin de s'adapter aux enjeux de justice représentationnelle soulevés par les algorithmes de demain. Dans les faits et supposant sa mise en œuvre, l'une des fonctions principales de la CARIA serait la modernisation constante de sa pertinence sur le plan de l'éthique de l'IA. Si nous considérons que le phénomène de la performativité algorithmique est inhérent au déploiement de l'IA et qu'il persistera, nous croyons que ses formes et ses effets futurs ne doivent pas être limités à cette théorisation actuelle étant donné le dynamisme intrinsèque de l'IA. De plus, le marché de la justice représentationnelle est également amené à se transformer puisque la frontière entre l'impérialisme culturel et les groupes y étant soumis est poreuse et fluctuante. L'exploitation de la performativité algorithmique se doit d'être dynamique, et c'est donc pour ces raisons qu'il est du devoir de la CARIA d'être constamment à l'affût des besoins de la justice sociale.

Conclusion

Nous positionnant dans le prolongement idéologique de *La Déclaration de Montréal*, la proposition de la mise sur pied de la Coalition arc-en-ciel représentative de l'intelligence artificielle établit manifestement que le déploiement de l'IA se doit d'être éthique et responsable. En ce sens, nous soutenons que l'articulation entre ces deux instances serait optimale dans l'atteinte d'une plus grande justice sociale. En supposant une attitude laxiste face au développement ainsi qu'à l'usage des SIA, nous affirmons, notamment en raison de l'émergence d'une force performative contraignante de la répétition des décisions algorithmiques, qu'une détérioration de la justice sociale surviendrait. En ce sens et en absence d'un cadre de développement responsable, nous concluons que le phénomène de la performativité algorithmique nous permet de démontrer que *l'utilisation grandissante des processus décisionnels automatisés engendre un accroissement des injustices structurelles*. Puisque nous avons déterminé que la force performative des algorithmes renforce l'oppression de l'impérialisme culturel à chaque fois qu'une injustice représentationnelle est produite ou reconduite par un biais algorithmique, nous n'avons d'autre choix que de déclarer que le déploiement non réglementé de l'IA entraînera une augmentation des injustices structurelles ainsi qu'une détérioration de la justice sociale. Si notre analyse du phénomène de la performativité algorithmique nous permet effectivement d'affirmer que la réglementation de l'IA est nécessaire pour combattre les injustices structurelles, elle nous permet également d'imaginer une société au sein de laquelle les SIA participeraient à l'édification d'une plus grande justice sociale. Afin de démontrer comment il était possible d'exploiter les processus décisionnels automatisés afin qu'ils deviennent socialement bénéfiques, nous avons présenté le fonctionnement de l'IA ainsi que ses implications, le phénomène de la performativité algorithmique et son effet sur la construction identitaire, le cadre de justice sociale mobilisé nous menant vers une norme de justice, le tout pour culminer vers création de la CARIA, laquelle nous permettrait effectivement d'y parvenir.

Rappelons les étapes de notre raisonnement. Dans le premier chapitre, nous avons présenté le fonctionnement de l'IA en expliquant que la donnée constitue le nouveau carburant de cette quatrième révolution industrielle. Le développement et la complexification des SIA déterministes et apprenants, qu'ils soient en apprentissage

machine ou en apprentissage profond, est nécessairement tributaire de l'accumulation exponentielle des données et de ce qu'elles rendent, et rendront possible pour le futur. Nous avons également démontré que les complexes algorithmiques, que nous avons qualifiés d'*eventful* étant donné le dynamisme intrinsèque de ces derniers, ont désormais la capacité de résoudre un nombre grandissant de problèmes et que c'est pour cette raison que les processus décisionnels automatisés se déploient dans toutes les sphères de la vie quotidienne. En ce sens, nous n'avons d'autres choix que d'accepter que les décisions algorithmiques qu'ils génèrent ont des implications concrètes dans la vie des individus. À partir de l'évidence des répercussions qu'elles engendrent, qui sont pour la plupart inintelligibles, nous avons conclu qu'il est impératif d'analyser concrètement leurs effets sur la population afin d'être en mesure de nous prononcer sur leur désirabilité. En ce sens, dans le deuxième chapitre, nous avons abordé la question de la théorisation de la performativité algorithmique, phénomène permettant d'expliquer concrètement l'effet de la force performative des algorithmes sur la construction identitaire des individus. En établissant que la quantité de décisions algorithmiques générées par les SIA est maintenant suffisante pour expliquer l'émergence d'une force performative des algorithmes, et en considérant que cette tendance à la hausse persistera, il devient possible d'anticiper la progression d'une convergence entre l'identité sociale et l'identité numérique. Étant toutes deux construites performativement, d'une part à l'intersection des appartenances sociales et d'autre part à l'intersection des classifications algorithmiques, nous avons précisé que la modernisation de la socialisation provoque effectivement une interaction performative entre ces dernières, laquelle se traduit ultimement par leur uniformisation en une identité réellement vécue et incarnée. Nous redoutons donc que les décisions algorithmiques déterminent performativement la construction identitaire en classifiant les individus selon des catégories, ce qui est potentiellement contraignant du point de vue de la norme de justice que nous avons établie par la suite. En effet, au cours du troisième chapitre, nous avons eu comme objectif de prouver, notamment grâce à la mesure de la norme de justice de la libre construction identitaire, que le phénomène de la performativité algorithmique contribue aux injustices structurelles. Pour ce faire, nous avons adopté un cadre théorique de justice sociale qui permettrait effectivement de traiter des enjeux soulevés par le déploiement de l'IA, et plus précisément, de ceux occasionnés par la force performative

des algorithmes. Nous avons d'abord tenté de traiter des injustices algorithmiques grâce au cadre neutraliste de Rawls, mais la conception de neutralité y étant promue rendait impossible leur détection. En contrepartie, le cadre théorique proposé par Young nous permettait de rendre compte de l'effet performatif des algorithmes sur la justice sociale. En ce sens, les notions de groupes sociaux, d'oppression et d'impérialisme culturel ont rendu possible l'analyse, ainsi que la compréhension, des mécanismes et processus expliquant le lien entre ce phénomène et les effets qu'il génère. Finalement, il nous aura été possible d'affirmer que la force performative des algorithmes participe à la détérioration de la justice sociale puisque la reconduction automatique des biais algorithmiques interfère avec la construction identitaire des individus. Dans un quatrième et dernier chapitre, nous avons proposé une piste de solutions à la problématique du renforcement de l'oppression de l'impérialisme culturel provoqué par le pouvoir performatif des décisions algorithmiques. Afin d'y parvenir, nous avons prouvé que le phénomène de la performativité algorithmique n'est pas fondamentalement injuste, mais qu'il le devient lorsque la reconduction des biais algorithmiques générés contribue à la dominance des groupes culturellement privilégiés. Inversement, nous estimons qu'il est théoriquement possible d'exploiter ce phénomène afin qu'il participe plutôt à l'atteinte d'une plus grande justice sociale puisqu'une manipulation planifiée des biais algorithmiques, impliquant nécessairement une réorientation de la force performative des décisions algorithmiques, se traduit par l'affaiblissement de l'impérialisme culturel ainsi que par une dissolution de la catégorisation contraignante des algorithmes. Ayant comme objectif principal de collaborer à la réglementation du déploiement de l'IA afin qu'il respecte les exigences de la justice sociale, et qu'il s'ajuste à la norme de justice de la non-interférence identitaire, nous avons proposé l'instauration de la CARIA. Le mandat de la coalition serait de diriger l'exploitation de la performativité algorithmique afin que la représentativité des groupes minoritaires soit culturellement normalisée. C'est justement grâce à la modélisation artificielle des biais algorithmiques que les experts de la CARIA pourraient s'assurer que la force performative des algorithmes participe positivement à l'émancipation des individus représentant les groupes minoritaires au lieu de contribuer négativement à leur stigmatisation contraignante.

Bien que la conclusion de notre mémoire confirme notre inquiétude de départ, à savoir celle de *l'accroissement des injustices structurelles généré par l'utilisation grandissante des processus décisionnels automatisés*, nous sommes néanmoins d'avis que cette éventualité ne constitue pas une fatalité étant donné le pouvoir décisionnel que l'être humain détient finalement en amont du déploiement de l'IA. Tout comme le soutient la Déclaration de Montréal, nous sommes également d'avis que l'encadrement responsable de l'IA est nécessaire puisqu'une attitude passive à son égard engendrerait inévitablement des retombées indésirables pour la société. Si la première étape pour freiner cette progression réside dans la prévention des injustices générées, nous soutenons de surcroît que la direction à suivre devrait être celle de l'exploitation de la performativité algorithmique puisque les exigences de la justice sociale le commandent. Véritablement, l'IA est un outil dont le potentiel s'est érigé en une arme à double tranchant pour l'humanité, et dont la nature des effets dépendra ultimement de l'usage que nous en ferons. Si nous sommes pour notre part optimistes par rapport au projet mis de l'avant par la Déclaration de Montréal, et que nous sommes également d'avis que ce dernier devrait s'étendre au-delà des frontières de la réglementation du développement de l'IA pour éventuellement s'immiscer au sein de son orientation, c'est principalement parce que nous jugeons que le stade embryonnaire du déploiement de l'IA nous permet encore une certaine marge de manœuvre. Nous jugeons que la société doit se doter des outils théoriques nécessaires afin d'être en mesure de saisir justement cette opportunité d'influencer positivement son développement.

A contrario, il serait possible d'argumenter que l'abolition de cette technologie est plus sécuritaire étant donné que sa maîtrise dépasse largement notre capacité collective. Effectivement, la recherche en éthique de l'IA est prolifique, mais elle tarde à suivre la cadence avec laquelle s'accumulent les nouvelles applications et SIA. De plus, les pouvoirs inégalés concentrés dans les mains des géants du Web font en sorte qu'ils s'opposent généralement à toutes formes d'ingérence dans leurs modèles, et il est clair que les organisations mandatées pour leur réglementation ne détiennent pas les moyens nécessaires pour imposer une vision de l'IA responsable. Même si nous abondons en ce sens, et que nous sommes conscients de l'asymétrie caractérisant les rapports de force entre les intérêts concernés, nous persistons tout de même à croire que la Déclaration de Montréal est

avantageée par le fait qu'elle représente les intérêts de la population. À titre d'exemple, la certification d'aliments biologiques n'a jamais été imposée aux géants de l'agriculture, mais s'est plutôt popularisée et légitimisée grâce aux exigences citoyennes. De la même manière, la certification des SIA responsables, et peut-être même socialement bienfaisants, pourrait potentiellement être réclamée de la part des consommateurs et cela influencerait assurément l'engagement futur des entreprises en IA. À vrai dire, ce scénario nous apparaît plus réalisable que celui de la dissolution orchestrée des nouvelles technologies de l'IA. Si la décroissance technologique peut être justifiable dans la préservation de la justice sociale, nous craignons néanmoins que sa mise en œuvre réelle soit impossible considérant l'importance du marché de l'IA au sein de l'économie mondiale. Entre la réglementation, la manipulation et l'abolition, il existe une multitude de paramètres qu'il est de notre devoir de structurer afin que soit préservée la justice sociale. Nous espérons donc que la théorisation de la performativité algorithmique contribuera aux efforts mobilisés dans l'orchestration du développement responsable de l'IA de demain.

Bibliographie

- ALLARD, André. *Muhammad Ibn Musa Al-Khwarizmi, Le calcul Indien (algorismus)*. Versions latines du XIIe siècle, Librairie scientifique et technique Albert Blanchard, Paris, 1992.
- ALLARD, Laurence, et Olivier BLONDEAU, « L'activisme contemporain : Défection, expressivisme, expérimentation », *Rue Descartes*, numéro 55, 2007.
- AUSTIN, J. L., *Quand dire, c'est faire*, (traduit de l'anglais et introduction par Gilles Lanes), Paris, Édition du Seuil, 1970 [1962].
- BERGER, Peter L. et Thomas LUCKMANN, *La construction sociale de la réalité*, (traduit de l'anglais Pierre Taminiaux), Armand Colin, 2014, [1966].
- BILGE, Sirma, *Théorisation féministe de l'intersectionnalité*, Presses Universitaires de France, 2009.
- BOYD, Dana, *Privacy and Publicity in the Context of Big Data*, 2010.
- BRUNEAULT, Frédéric et al. *Former à l'éthique de l'IA en enseignement supérieur : Référentiel de compétence*, 2022.
- BUCHER, Tania. « Neither black nor Box », dans BUCHER, Tania (dir.), *If...Then: Algorithmic Power and Politics*, Oxford University Press, New-York, 2018.
- BUTLER, Judith, *Trouble dans le genre, Le féminisme et la subversion de l'identité* (traduit de l'anglais par C. Kraus), 1990.
- CARDON, Dominique, « Le design de la visibilité, Un essai de cartographie du web 2.0 », *Réseaux*, numéro 152, 2008.
- DATTA, Amit, Michael Carl TSCHANTZ, et Anupam DATTA, « Automated Experiments on Ad Privacy Settings », *Proceedings on Privacy Enhancing Technologies*, 2015.
- DE BEAUVOIR, Simone, *Le deuxième sexe, Les faits et mythes*, Paris, Gallimard, 1949.
- Déclaration de Montréal IA responsable, *La déclaration de Montréal pour un Développement responsable de l'Intelligence Artificielle*, 2018.
- DENIS, Jérôme. « Préface : Les nouveaux visages de la performativité », *Études de communication : Langages, information, médiations*, OpenEdition Journals, numéro 29, 2006.
- DESMONS, Ophélie, *Les présupposés du libéralisme politique : quelle justification?*, 2014.
- DWORKIN, Ronald. *Taking rights seriously*, 1977.
- EAGLE, Gill, « Learning to Become a "Natural Woman": The Process of Socialisation », *Empowering Women for Gender Equity*, no 2, 1988.
- ERTZCHEID, Olivier, *Qu'est-ce que l'identité numérique? Identité numérique et e-réputation*, OpenEdition Press, Marseille, 2013.
- EUBANKS, Virginia, *Automating Inequality, How High-Tech tools profile, police, and punish the poor*, St. Martin's Press, New-York, 2018.
- FERGUSON, Andrew, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, 2017.
- GOFFMAN, Erving, *The presentation of the self in everyday life*, Écosse, 1956.

- HICKS, Marie, *Programmed Inequality, How Britain discarded women technologist and lost its edge in computing*, 2017.
- JOHNSTON, Leslie, «A “Library of Congress” Worth of Data: It’s All In How You Define It», *Library of Congress*, 2012. En ligne au : <https://blogs.loc.gov/thesignal/2012/04/a-library-of-congress-worth-of-data-its-all-in-how-you-define-it/#:~:text=You%20Define%20It-.A%20%E2%80%9CLibrary%20of%20Congress%E2%80%9D%20Worth%20of%20Data%3A%20It's%20All%20In%20How%20You%20Define%20It&text=That's%20the%20equivalent%20of%2018,Library%20of%20Congress%20today.%E2%80%9D%20LINK>, consulté le 19 juillet 2022.
- Journal officiel de l’Union européenne, *Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l’égard du traitement des données à caractère personnel et à la libre circulation des ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données)*, 2016.
- KEARNS, Micheal et Aaron ROTH. *The ethical algorithm: the science of socially aware algorithm design*, Oxford University Press, New-York, 2019.
- KLEYMAN, Bill, « The big Data Battleground: Analyzing the Big Picture », *DataCenter Knowledge*, mis à jour le 12 septembre 2012. En ligne au : <https://www.datacenterknowledge.com/archives/2012/09/12/a-look-into-the-big-data-battleground-analyzing-the-market?utm-source=feedburner&utm-medium=feed&utm-campaign=Feed%3A+DataCenterKnowledge+%28Data+Center+Knowledge%29>, consulté le 8 octobre 2020.
- LAHLOU, Saadi, « Identity, social status, privacy and face-keeping in digital society », *Social Science Information: Cognitive technologies*, 2008.
- LARMORE, Charles, *Modernité et morale*, 1993.
- LARSON, Jeff, MATTU, Surya, KIRCHNER, Lauren et Julian ANGIN, « How We Analyzed the COMPAS Recidivism Algorithm », *ProPublica*, mis à jour le 23 mai 2016. En ligne au : <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>, consulté le 5 septembre 2020.
- LEWES, G. H. *Problem of Life and Mind*, The Riverside Press, Cambridge, 1875.
- MÉGNY, Mael et Issam IBNOUHSEIN. *Quelle transparence pour les algorithmes d’apprentissage machine?* Archive ouverte pluridisciplinaire, 2018.
- MEHRABI, Ninareh, et AL, *A survey on Bias and Fairness in Machine Learning*, 2019.
- NÉRON, Pierre-Yves, *Rethinking the very idea of egalitarian markets and corporation*, 2015.
- NOBLE, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*, New-York University Press, New-York, 2018.
- OHM, Paul. « Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization », *UCLA Law Review*, vol. 53, 2010.
- PASQUALE, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge, 2015.

- PATEL, Keyur et Sunil PATEL. « Internet of things – IOT : Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges », *International Journal of Engineering Science and Computing*, vol. 6, n°5, May 2016.
- PERES, Éric. *Les données numériques, un enjeu d'éducation et de citoyenneté*, Paris, Les éditions des journaux officiels, 2015.
- PESCARU, Maria, *The importance of the socialization process for the integration of the child in the society*, University of Pitesti, 2019.
- RAHMAN, Farhan, « COMPAS Case Study: Fairness of a Machine Learning Model », *Towards Data Science*, mis à jour le 7 septembre 2017. En ligne au : <https://towardsdatascience.com/compas-case-study-fairness-of-a-machine-learning-model-f0f804108751>, consulté le 5 septembre 2020.
- RAWLS, John, *Théorie de la justice* (Traduit par Catherine Audard), Éditions Points, 2009, [1971].
- RAWLS, John, *Libéralisme politique* (Traduit par Catherine Audard), PUF, 1995.
- SARTRE, Jean-Paul, *L'existentialisme est un humanisme*, Paris, Éditions Nagel, 1946.
- SCHATZKI, Theodore R., Karin Knorr CETINA et Eike Von SAVIGNY, *Practice turn in contemporary theory*, London, 2001.
- SCHECHNER, Richard, *Performance studies: An introduction*, New-York, 2002.
- SEJNOWSKI, Terrence J. *The Deep Learning Revolution: Artificial intelligence meets human intelligence*, The MIT Press, Cambridge, 2018.
- SHAKESPEARE, William, *As you like it*, Londres, 1599.
- SURESH, Harini et John V. GUTTAG, *A framework for understanding unintended consequences of Machine learning*, 2020.
- SWEENEY, Latanya. *Simple Demographics often Identify People Uniquely*, Carnegie Mellon University, Pittsburgh, 2000.
- TURING, Alan, *Computing Machinery and Intelligence*, 1950.
- VILLANI, Cédric, *Donner un sens à l'intelligence artificielle, Pour une stratégie nationale et européenne*, 2018.
- WARNER, Leah R. et Stephanie A. SHIELDS, « The intersections of Sexuality, Gender, and Race: Identity Research at the Crossroads », *Sex Roles*, Vol. 68, New-York, 2013.
- YOUNG, Iris Marion, *Justice dans the politics of difference*, Princeton classics, 1990.
- ZAFFAGNI, Marc, « Cette IA a appris à conduire une voiture autonome en 20 minutes », Futura Sciences, mis à jour le 7 septembre 2018. En ligne au : <https://www.futura-sciences.com/tech/actualites/intelligence-artificielle-cette-ia-appris-conduire-voiture-autonome-20-minutes-71942/>, consulté de 22 juillet 2020.
- ZIMMER, Mathieu. *Apprentissage par renforcement développemental*, Archive ouverte pluridisciplinaire, 2018.