Discrimination à l'ère des algorithmes : rouages et enjeux sociétaux

ANH LE Duc **lub.the.studio@gmail.com**
Master 1 Optimisation et Recherche Opérationnelle

DIALLO Mamadou **diallocire371@gmail.com**
Master 1 Apprentissage et Traitement Automatique de la Langue

HALIMI Khedaoudj **jupia115@gmail.com**
Master 1 Optimisation et Recherche Opérationnelle

MERCIER Juanfer **ercier.juanfer@gmail.com**
Master 1 Optimisation et Recherche Opérationnelle

I. Résumé :

L'intégration et le déploiement des algorithmes sont en forte évolution, ayant atteint le statut de technologie et convoités pour leur puissance, ils sont largement présents dans de nombreux domaines : aide à la décision, véhicules autonomes, mais aussi touchent de plus en plus de nouveaux secteurs : éducation, système pénitencier. S'ils sont présentés en général comme des éléments objectifs et véridiques, il n'en demeure pas des éléments contestés et contestables. Leur mécanisme de décision conduit généralement à des pratiques discriminatoires sur le genre, la race ou l'ethnie. L'objectif de cet article est de décrire les discriminations liées aux algorithmes, les mécanismes qui entrainent cet état de fait, les enjeux sociétaux, les conséquences engendrées et menaces que ces éléments peuvent représenter. Enfin, une présentation de quelques pistes sera faite pour remédier à cet état de fait.

II. Introduction :

« Algorithme », ce mot qui peut se définir comme une suite d'instructions permettant de résoudre un problème bien défini, a depuis son apparition, attiré l'attention de tout genre : industriels, scientifiques, étudiants pour ne citer que ceux-ci, suscitant beaucoup l'engouement. L'idée de l'intelligence artificielle a vu le jour lorsque Alan Turing s'est posé la question à savoir « les machines peuvent-elles penser ? », et a publié dans la foulée son article ***Computing Machinery and Intelligence***. Mais c'est en 1956 dans un département de Mathématiques à Darthmouth que l'intelligence artificielle est devenue une science [1][1], où un groupe soutint que l'intelligence pouvait être démontré par la capacité à jouer à des jeux et particulièrement aux échecs. L'exemple de Kasparov qui était considéré comme le meilleur joueur perdant face à la machine Deep Blue de IBM restera toujours dans les annales. Mais l'intelligence est bien plus complexe que sa réduction en la capacité à jouer aux jeux ou aux échecs en particulier, les facultés ou les capacités d'apprentissage de l'humain sont influencées par plusieurs aspects (environnementaux, sociétaux et même physiques) qui peuvent ne pas pouvoir être représentés ou codifiés algorithmiquement. Dans le monde informatique, un groupe de personnes affirmera qu'il y a deux façons de programmer les ordinateurs, un autre groupe de personnes affirmera qu'il y a deux façons de donner de la connaissance à un ordinateur. En dépit de la différence des expressions, ces deux groupes, pointent le fait qu'il peut soit être fourni à l'ordinateur une suite d'instructions dans un langage de programmation reconnu par l'ordinateur : c'est-à-dire

[1] https://fr.wikipedia.org/wiki/Intelligence_artificielle#Historique

de prendre toutes les connaissances du monde et les mettre dans l'ordinateur ; soit donner à l'ordinateur les capacités de l'intelligence humaine : c'est-à-dire avoir la capacité d'apprendre et laisser l'ordinateur découvrir l'écosystème mondiale afin de se doter de sa propre intelligence à travers les éléments qu'il aura récupéré de cet écosystème qu'il transformera en connaissance. Cette dernière méthode est appelée apprentissage machine en anglais « Machine Learning »[2]. L'apprentissage machine, est un champ d'étude de l'intelligence artificielle n'ayant été popularisé que récemment à travers l'utilisation croissante du big data[3], et est à la base de développement d'outils mathématiques et informatiques pour faire des prédictions, ou prendre des décisions sur la base de modèles construits par les algorithmes d'apprentissage machine. Ces algorithmes sont décisionnaires, deviennent largement répandu, et sont présentés comme des éléments objectifs et véridiques car loin des décisions d'humeurs ou imparfaites humaines et sont basés seulement sur des données, des chiffres, des vérités mathématiques ; mais ils sont dénués généralement de toute éthique. Qu'il s'agisse d'aider à déterminer qui est embauché, licencié, qui obtient un prêt, ou combien de temps un individu passe en prison, les décisions qui étaient traditionnellement prises par des humains sont rapidement prises par des algorithmes (O'Neil, 2017, Citron and Pasquale, 2014). Les algorithmes d'apprentissage machine sont entrainés avec des exemples de données, et il a été montré que les algorithmes entrainés avec des données (appelé données étiquetés) biaisées donnent lieu à une discrimination algorithmique (Bolukbasi et al.[2], 2016 ; Caliskan et al., 2017[3]). Bulukbasi et al. ont même montré que le modèle Word2Vec utilisé pour le prolongement lexical a en son sein des biais de genre : les auteurs ont utilisé Word2Vec pour entrainer un générateur analogique qui remplit les mots manquants dans les analogies. L'analogie « Man is to Computer Programmer as Woman is to … » a été compléter par le mot anglais « Homemaker » se conformant au stéréotype selon lequel la programmation est associée aux hommes et les tâches ménagères aux femmes. Ce modèle étant très utilisé, ses imperfections et ses biais seront propagés dans tout autre système l'utilisant. Il peut être dit que la machine n'a fait que retourner les exemples de discriminations ou de biais existant déjà dans la société. Comme exemple, le chatbot d'intelligence artificielle Tay.ai de Microsoft, qui a été désactivé en moins de 24h [4][4]. Le chatbot était une intelligence artificielle dont le but était d'étudier la compréhension du langage, même si disposant préalablement de réponses toutes faites pour certains sujets, il était créé pour apprendre à partir des interactions humaines, mais était devenu en 16h un programme raciste, misogyne [5][5].

III. Discrimination :

Les enjeux sociétaux face à l'évolution et l'expansion des algorithmes sont énormes et il arrive de s'interroger si ces enjeux diffèrent d'une société à une autre ou d'un Etat à un autre. Les exemples documentés et la recherche sur les sujets de discriminations et d'inégalités en

[2] Machine Learning : en français apprentissage machine « est l'étude scientifique des algorithmes et des modèles statistiques que les systèmes d'ordinateurs utilisent pour améliorer progressivement leur performance sur une tâche spécifique » [Wikipedia [en]]

[3] Big data : données massives. Le terme désigne la quantité gigantesque de données générées chaque jour par des utilisateurs du monde entier

[4] M.J. Wolf, K.W. Miller, F.S. Grodzinsky : Why We Should Have Seen That Coming: Comments on Microsoft's Tay "Experiment," and Wider Implications,

[5] Fuchs, Daniel J.. 2018. "The Dangers of Human-Like Bias in Machine-Learning Algorithms." Missouri S&T's Peer to Peer 2, (1)

intelligence artificielle existent, mais ils sont en grande majorité le fait de chercheurs anglo-saxons et prennent comme cas concrets d'étude des algorithmes mis en œuvre aux Etats-Unis[6][6]. Lorsqu'un algorithme est mis en place et déployé, le seul retour espéré est qu'il effectue le travail pour lequel il a été mis en place, et il peut être possible de s'interroger comment l'algorithme arrive à sa conclusion, et pour les algorithmes dit de scoring[7], comment l'algorithme définit un être humain. Le choix des mesures qui traduisent la perception des ingénieurs dans la mise en œuvre des algorithmes peuvent souvent amener l'algorithme à prendre des décisions biaisées, car les ingénieurs sont humains et les humains ont des biais. Il existe des recherches en psychologie et sciences cognitives qui montrent l'existence de biais cognitifs dans la prise de décision (Khaneman et Tversky, 1974). Il peut être cité les biais de stéréotype qui sont un ensemble de croyance sur les caractéristiques typiques des membres d'un groupe social ou ethnique, les biais de confirmation qui consiste à privilégier les informations qui confirment ses hypothèses et négliger les autres informations, et l'effet de « bandwagon » ou d'« entrainement » qui consiste à imiter ou rejoindre la majorité, sans aucun autre critère ; qui peut conduire les ingénieurs à réutiliser des modèles populaires sans s'assurer de leur exactitude. L'ensemble de ces biais peuvent conduire à des choix de mesures qui traduisent la vision d'un groupe de personnes, et qui vont entrainer l'algorithme à prendre des décisions biaisées. Ces systèmes déployés à large échelle conduisent à des discriminations, et renforcent les biais de confirmation, de la même façon que les données utilisées pour faire apprendre ces algorithmes ont été collecté de façon approximative ou y ont été introduit des corrélations fallacieuses.

1. **Système scolaire :**

Il peut être possible de s'interroger sur le processus par lequel les étudiants sont orientés dans les Universités, ou encore de s'interroger comment les dossiers de candidatures dans un master sont traités. La forte évolution de l'intelligence artificielle et des algorithmes d'apprentissage automatique, a mis à la mode l'automatisation de l'administration pour le traitement des dossiers. Mais au-delà des étudiants les enseignants aussi sont soumis à des systèmes de notation en fonction de leur performance. Dans le livre « Weapon of Math Destruction »[7], un exemple concret est le cas du projet IMPACT aux Etats-Unis, qui est un « modèle de valeur ajoutée »[8]. Le lancement du projet visait à évaluer les compétences des professeurs dans l'enseignement de sa matière dans les écoles publiques du district de Washington. L'évaluation portait sur une note donnée par un algorithme, et si cette note était en dessous d'un certain seuil, l'enseignant était qualifié comme « incompétent » et devait être renvoyé. En dépit de la méconnaissance des éléments pris en compte dans la mesure, et du manque de rigueur statistique en raison de la taille de l'échantillon, cette note prévalait sur les appréciations de l'administration, des parents d'élèves et même de la collectivité. Ainsi par cette note des enseignants compétents se virent mis à la porte sans aucune possibilité de recours. Mais quel

[6] Patrice Bertail, David Bounie, Stephan Clémençon et Patrick Waelbroeck : Algorithmes : biais, discrimination et équité

[7] Scoring : est une technique mettre en place un système de prédiction pour générer des valeurs représentant un score de probabilité.

[8] Le modèle de valeur ajoutée appelé « value-added model » est un modèle appliqué aux Etats Unis. En comparant les résultats scolaires d'un groupe d'élèves d'un même niveau d'une année à autre, les progrès d'un professeur sont évalués dans l'enseignement d'une matière donnée [Cathy O'Neil : Weapon of Math Destruction]

lien avec la discrimination ? L'enseignement public, donc des gens de la classe moyenne ou pauvre se voient ainsi priver d'enseignant de qualité pour la plupart alors que l'enseignement privé et riche, inaccessible à la classe moyenne, ne tiennent absolument pas compte de cette évaluation. Un autre aspect de discrimination est le manque de possibilité d'avoir un recours ou d'avoir une explication sur le processus de décision.

2. Reconnaissance faciale :

L'accroissement des systèmes de reconnaissance faciale et leur intégration sont en forte évolution, et s'ils sont généralement présentés comme des outils de sécurité, il n'en demeure pas moins intrusif et plus inquiétant, imprécis. La précision des modèles est très importante. Big Brother Watch[9], dans une étude concernant l'utilisation de la reconnaissance faciale par la police britannique, déclare que leur logiciel était à 93% imprécis [10][8]. Le logiciel avait identifié, généralement des personnes issues de la communauté noire ou arabe, comme appartenant à des groupes terroristes, avec une confiance de correspondance supérieure à 93%, ce qui a entrainé des interpellations et des fouilles. Il est possible de se demander quel est le cadre légal, généralement aucun. En France au-delà de la loi informatique et libertés (1978) et le RGPD[11] (2018) qui posent les principes du cadre légal, il y a la proposition de loi « d'expérimentation créant un cadre d'analyse scientifique et une consultation citoyenne sur les dispositifs de reconnaissance faciale par l'intelligence artificielle »[12][9], qui vise à créer un cadre d'expérimentation transparent et éthique pour les technologies de reconnaissance faciale par l'intelligence artificielle. Cette démarche vise à garantir un usage responsable de ces technologies. Aux Etats-Unis, plus de 117 millions de personnes ont leur visage dans un réseau de reconnaissance faciale utilisable par la police, sans véritable loi de protection, ni au niveau fédéral, ni au niveau des Etats[1314][10][11]. L'époque est à l'ère de l'automation des données massives, et les grands groupes, géants de la technologie ont accès à nos données et les utilisent dans divers cas. Meta[15] avait déposé un brevet[1617][12][13] pour permettre le payement dans les magasins, à travers la reconnaissance faciale, et par la même occasion, donner une note de confiance aux clients. En plus du fait que leur système de reconnaissance faciale et d'identification n'est pas correcte : leur système de reconnaissance Face++, d'après une étude[18][14], montrait un taux général de 16.5% sur les personnes de couleurs noirs et individuellement 34.5% sur les femmes noires ; ce système créera une société de surveillance où les droits des personnes se résumeront à une note de confiance, attribué par une société, et dont personne ne connait le système de notation. Ces systèmes seront des éléments qui catégorisent et fichent les individus, dans le but de maintenir un ordre social et déterminer qui

[9] Big Brother Watch est une organisation britannique à but non lucratif pour les libertés civiles et la vie privée fondé en 2009 [Wikipedia [en]]
[10] Big Brother Watch : Big Brother Watch Briefing on facial recognition surveillance
[11] RGPD : Règlement Général sur la Protection des Données
[12] https://www.assemblee-nationale.fr/dyn/15/textes/l15b4127_proposition-loi#
[13] https://www.perpetuallineup.org/
[14] https://medium.com/mit-media-lab/the-algorithmic-justice-league-3cc4131c5148
[15] Méta est le nouveau nom de l'entreprise Facebook
[16] https://www.biometricupdate.com/201711/facebook-files-patent-for-facial-recognition-for-physical-payments
[17] https://www.pymnts.com/facebook/2017/facebook-patent-facial-recognition-authentication/
[18] Gender Shades : Intersectional accuracy disparities in commercial gender classification

peut faire ou pas une activité spécifique. Et pour certains pays, traquer les dissidents politiques. Les performances d'un système de reconnaissance faciale peuvent être présentées de manière agrégées, mais, lorsqu'une décomposition des résultats du système par sous-groupe est faite, les performances diffèrent considérablement[19][15].

Dans le tableau ci-dessous, se trouve les performances de la classification de genre, mesurée par la valeur prédictive positive (VPP), le taux d'erreur (1 - VPP), le taux de vrais positifs (TVP) et le taux de faux positifs (TFP) des trois classificateurs commerciaux, évalués sur l'ensemble des données des référentiels des parlements pilotes « Pilot Parliaments Benchmark (PPB) ».

Classifieur	Métriques (%)	Tous	femmes	hommes	Peau foncée	Peau claire	femmes foncées	hommes foncés	femmes claires	hommes clairs
MSFT	VPP	93.7	89.3	97.4	87.1	99.3	79.2	94.0	98.3	**100**
	Taux d'erreur	6.3	10.7	2.6	12.9	0.7	**20.8**	6.0	1.7	0.0
	TVP	93.7	96.5	91.7	87.1	99.3	92.1	83.7	**100**	98.7
	TFP	6.3	8.3	3.5	12.9	0.7	**16.3**	7.9	1.3	0.0
Face++	VPP	90.0	78.7	99.3	83.5	95.3	65.5	**99.3**	94.0	99.2
	Taux d'erreur	10.0	21.3	0.7	16.5	4.7	**34.5**	0.7	6.0	0.8
	TVP	90.0	98.9	85.1	83.5	95.3	98.8	76.6	**98.9**	92.9
	TFP	10.0	14.9	1.1	16.5	4.7	**23.4**	1.2	7.1	1.1
IBM	VPP	87.9	79.7	94.4	77.6	96.8	65.3	88.0	92.9	99.7
	Taux d'erreur	12.1	20.3	5.6	22.4	3.2	**34.7**	12.0	7.1	0.3
	TVP	87.9	92.1	85.2	77.6	96.8	82.3	74.8	**99.6**	94.8
	TFP	12.1	14.8	7.9	22.4	3.2	**25.2**	17.7	5.20	0.4

Tableau 1 : source [Gender Shades : Intersectional Accuracy Disparities in Commerial Gender Classification (Tableau numéro 4)].
Les classifieurs du tableau présentent les taux d'erreur les plus élevés pour les femmes à la peau foncée (de 20,8% pour Microsoft à 34.7% pour IBM).

3. Commerce et publicité en ligne : l'industrie du clic :

La rentabilité des algorithmes dans le ciblage de la publicité en ligne a rapporté respectivement pour Meta et Google, 114.93 milliards[20][16] de dollars U.S. et 209.49 milliards[21][17] de dollars

[19] Gender Shades : Intersectional Accuracy Disparities in Commerial Gender Classification
[20] https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide/
[21] https://www.statista.com/statistics/266249/advertising-revenue-of-google/

U.S. pour l'année 2021. Les algorithmes d'apprentissage machine utilisent des données passées pour prédire le futur[22][18]. Que ce soit explicite ou non explicite, les utilisateurs des espaces de vente en ligne ont, un profil type associé à leur personne. Lorsqu'un utilisateur fait défiler son fil d'actualité, ou effectue des recherches sur un moteur de recherche, une bataille algorithmique est enclenchée pour lui montrer le prochain post qui peut attirer son attention. Quoi de plus efficace que l'industrie des données massives qui collecte, et traite les données utilisateurs, dévoilées volontairement. Ce faisant les utilisateurs sont devenu des sujets d'études grâce aux données partagées, et ces études prennent en compte plusieurs paramètres comme l'âge, le sexe, l'environnement social, etc. Ces paramètres permettent d'étudier entre autres le pouvoir d'achat des utilisateurs. Les algorithmes dans leur choix, sont intentionnellement ou inintentionnellement manipulés, pour toucher une plus grande audience ou un groupe visé d'une plateforme. Une étude[23][19] menée par des chercheurs de l'Université Northeastern associés à l'entreprise Upturn, sur l'algorithme publicitaire de Meta, a montré que le groupe d'utilisateurs à qui l'algorithme fait le choix de montrer des publicités peut être biaisé en fonction du sexe ou de la race. Les chercheurs ont pu remarquer que l'algorithme de Meta se concentrait sur des publics spécifiques pour les publicités particulières mises en place dans le cadre de l'expérience : 75% de noirs pour les offres d'emploi pour conducteur de taxi, 85% de femmes pour les offres d'emploi de caissier de supermarché, 72% de blancs dont 90% d'hommes pour les emplois de bûcheron, et un poste en intelligence artificielle avait touché 47% d'utilisateurs issus des minorités contre 64% pour un travail de concierge. Lambrecht et Tucker (2018) [20] ont étudié, par exemple, comment un algorithme fournissant des annonces publicitaires faisant la promotion d'emplois dans les domaines des sciences, de la technologie, de l'ingénierie et des mathématiques (STEM) peut discriminer les femmes. Les auteurs ont montré qu'un algorithme qui optimise simplement le rapport coût-efficacité de la diffusion d'annonces affiche moins d'annonces destinées aux femmes, car le prix du segment des jeunes femmes est supérieur à celui des jeunes hommes [21]. Si ces éléments montrent les conséquences sur l'accès équitable à une offre d'emploi, il y a d'autres secteurs où ce contrôle de la distribution de l'information peuvent avoir des conséquences négatives importantes pour la société : l'accès au crédit ou un logement.

4. Système pénal :

D'après l'American Civil Liberties Union[24] les peines infligées aux hommes noirs dans le système fédéral sont près de 20% plus longues que pour les blancs convaincus de crimes similaires. Une étude de l'université de Maryland dans le comté Harris à Huston, les procureurs étaient trois fois plus enclins à réclamer la peine de mort pour des prévenus noirs, et quatre fois plus pour les hispaniques que pour les blancs déclarés coupables des mêmes faits. Aux Etats-Unis, des travaux ont mis en évidence que les populations afro-americaines étaient plus souvent pénalisées par les décisions de justice qui s'appuient sur le recours aux algorithmes (Angwin et al. 2016)[22]. ProPublica[25] a évalué l'un des algorithmes de la société Northpointe COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)[26][23]. L'organisation

[22] Patrice Bertail, David Bounie, Stephan Clémençon et Patrick Waelbroeck : Algorithmes : biais, discrimination et équité
[23] https://arxiv.org/abs/1904.02095
[24] Union Américaine pour les libertés civiles
[25] ProPublica est une organisation à but non lucrative basé à New York
[26] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

a comparé sur l'ensemble d'un comté pendant deux ans, les taux de risque de récidive prédits par l'algorithme à ceux réellement observés lorsque les délinquants étaient remis en liberté. L'algorithme prédit correctement le risque de récidive dans 61% des cas : 59% des cas pour les afro-américains, et 63% des cas pour les blancs. Cependant, lorsque l'algorithme se trompe, il se trompe plus fréquemment pour les afro-américains que pour les blancs : les accusés blancs sont souvent prédits moins risqués qu'ils ne le sont et les accusés noirs, sont souvent prédits plus risqués qu'ils ne le sont. Il a été remarqué que les accusés blancs qui avaient récidivé dans les deux ans avaient été considérés à tort comme à faible risque, presque deux fois plus souvent que les récidivistes noirs (48% contre 28%). Et les accusés noirs qui n'ont pas récidivé dans les 2 ans sont plus fréquemment classés à tort dans la classe risque élevé que les blancs (45% contre 23%) [27][24]. Il peut donc être dit que l'algorithme surévalue le risque récidive des afro-américains, et sous-estime ce même risque pour les blancs.

IV. **Les algorithmes au service de l'humain** :

Il a été présenté comment les algorithmes dans leur mode discriminatoire, peuvent être dangereux et destructeurs mais, et si les algorithmes d'intelligence artificielle étaient au service de l'humain, dans une transparence absolue. Prenons en exemple les statistiques sportives, particulièrement dans le basket : une équipe ou un joueur est présenté par un ensemble de paramètres bruts pour le joueur représentant ses performances en saison régulière, en playoff[28] et même lors des All Stars game : nombre de minutes jouées, pourcentage de trois points tentées, pourcentage de trois points réussis, pourcentage de lancer franc et bien d'autres. Toutes ces valeurs sont fournies avec des mesures exactes, aucune donnée de substitution n'entre en jeu. Il est connu des fans de sport lorsqu'un match important est programmé les premiers éléments analysés sont les statistiques des joueurs individuellement afin de trouver un moyen de les arrêter, autrement dit comment ces chiffres entrent en ligne pour assurer une victoire à l'équipe ou quels joueurs réunir dans le club afin de maximiser ses chances d'être sacré champion de l'année. Des algorithmes d'analyses et de prédictions sont ainsi crées mais des algorithmes justes, transparents et accessibles par tous sans aucune distinction ainsi chaque parieur peut librement consulter les statistiques et décider lui-même quel pari l'apportera plus de gain et chaque équipe peut privilégier le paramètre qui l'intéresse le plus. Autre caractéristique de ces algorithmes c'est la capacité à être challengé régulièrement, car chaque saison de nouvelles statistiques arrivent et défient les prédictions faites par les algorithmes qui pourront être réajustés. Les algorithmes doivent suivre le même modèle que les algorithmes de prédictions sportives : être confronté à la réalité et être régulièrement mis à jour.

V. **Quelques pistes pour limiter biais :**

Les méthodes de récoltes des données ont grandement changé depuis l'avènement des données massives, et la réussite des algorithmes se mesure par des résultats chiffrés : profits rapportés aux entreprises, etc. Si l'aspect explicabilité des algorithmes d'apprentissage machine est un défi technique, leur mise à l'épreuve de la réalité peut être un bon début. Cette technique créera une boucle d'apprentissage de l'algorithme, qui se perfectionnera, en remettant en question son modèle face à la réalité des données nouvelles. Dans le rapport de Cédric Villani « Donner un

[27] https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
[28] Playoffs : est une compétition en série éliminatoire qui se déroule après la saison régulière et regroupent seize équipes

sens à l'intelligence Artificielle : Pour une stratégie nationale et européenne »[29][25], il met en lumière ce besoin d'explicabilité, aujourd'hui défi scientifique, qui consiste à « ouvrir la boite noire ». Il peut être cité comme exemple, le programme « Explanaible AI »[30][26] de la DARPA[31], qui a lancé un appel à proposition destiné à soutenir la recherche sur l'explicabilité de l'intelligence artificielle. Ce projet soutient trois axes : la production de modèles explicables, la production d'interfaces utilisateurs plus intelligibles et la compréhension des mécanismes cognitifs à l'œuvre pour produire une explication satisfaisante.

Les données recueillies ne doivent pas faire l'objet d'approximation, elles doivent être faire l'objet d'exactitudes, et recueillies par des méthodes statistiques établies et documentées. L'impact d'un algorithme dépend de l'échelle à laquelle il a été déployé, il doit exister un système de recours contre les décisions prises par les algorithmes. Ce système empêchera les algorithmes de restreindre l'accès à des éléments essentiels comme la recherche d'emploi ou l'accès à un crédit.

Pour limiter les discriminations issues des biais, il faudrait assurer la diversité des équipes de conception, de mise en œuvre et de déploiement des algorithmes : il est important que les acteurs (utilisateurs, responsables, etc.) soient de plus en plus impliqués dans la mise en place de ces algorithmes.

Bien que certaines discriminations soient involontaires : les algorithmes peuvent classifiés les utilisateurs en fonction des produits achetés et de ce fait peut différencier hommes et femmes ; les concepteurs doivent avoir une responsabilité dans les préjudices causés par les systèmes.

VI. Conclusion :

Le monde est dans l'ère des données massives, l'époque est ainsi faite et les algorithmes de décisions prennent plus de place dans le quotidien des humains, qui en grande majorité sont consciemment ou inconsciemment victimes de ces algorithmes par leurs décisions, qui sont sans appel et ne souffrent d'aucune remise en cause. De ce fait les décisions destructrices prises par les algorithmes peuvent produire des résultats erronés et donc accroitre des discriminations. Ces discriminations peuvent être basées sur le genre, la race ou l'ethnie et sont un danger pour la démocratie. Les problématiques sont donc réels et les enjeux sociétaux sont colossaux. L'équité, l'explicabilité des résultats, la régulation, entre autres, doivent être au cœur de la mise en place de ces systèmes dans une époque où social et technique sont difficiles à séparer.

Les algorithmes d'intelligence artificielle et d'apprentissage machine suscitent beaucoup l'engouement et gagnent plus de secteurs. Lorsque les sujets d'intelligence artificielle et d'apprentissage machine sont évoqués, les possibilités sont grandes, et les responsabilités nombreuses. Une responsabilité éthique et déontologique doit être imposé aux personnes, qui seront en avant dans la mise en place de ces outils. La prise en compte de ces aspects est importante, afin de non pas apporter une dimension humaine ou de compassion, mais que les éléments pris en compte dans leur mise en place soient justes, et aussi régulièrement remis en cause à l'image des algorithmes sportifs. Non pas que les modèles sportifs sont parfaits, mais

[29] https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf
[30] https://www.darpa.mil/program/explainable-artificial-intelligence#_blank
[31] DARPA : Defense Advanced Research Projects Agency

donnent un bon exemple de libre d'accès, de facilité de compréhension mais surtout de remise en cause et d'équité.

Il est important d'élargir cette réflexion, au-delà des pistes formulées dans la partie précédente, en incluant les entreprises, les chercheurs, le monde politique et les acteurs de la société, pour définir un cadre légal et réfléchir conjointement aux meilleures manières d'aborder les discriminations, l'équité et la réduction de biais dans les algorithmes.

Références :

[1] Wikipedia. "Intelligence artificielle " https://fr.wikipedia.org/wiki/Intelligence_artificielle#Historique , consulté le 07/03/2022

[2] Tolga Bulokbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, Adam Kalai. "Man is to Computer Programmer as woman is to Homemaker ? Debiasing word embeddings" July 21, 2016. https://arxiv.org/abs/1607.06520 , consulté, le 07/03/2022

[3] Aylin Caliskan, Joanna J. Bryson, Arvind Narayanan "Semantics derived automatically from language corpora contain human-like biases". Last revised May 25, 2017. https://arxiv.org/abs/1608.07187 , consulté le 07/03/2022

[4] M.J. Wolf, K.W. Miller, F.S. Grodzinsky. "Why We Should Have Seen That Coming: Comments on Microsoft's Tay "Experiment," and Wider Implications ". The ORBIT Journal. Volume 1, Issue 2. 2017, Pages 1-12. https://www.sciencedirect.com/science/article/pii/S2515856220300493 , consulté le 07/03/2022

[5] Fuchs, Daniel J.. 2018. "The Dangers of Human-Like Bias in Machine-Learning Algorithms." Missouri S&T's Peer to Peer 2, Volume 2, Issue 1. May 2018 https://core.ac.uk/download/pdf/229121681.pdf , consulté, le 07/03/2022

[6][18][21] Patrice Bertail, David Bounie, Stephan Clémençon et Patrick Waelbroeck. "Algorithmes : biais, discrimination et équité". Télécom ParisTech - 14Février 2019. https://www.telecom-paris.fr/wp-content-EvDsK19/uploads/2019/02/Algorithmes-Biais-discrimination-equite.pdf , consulté, le 07/03/2022

[7] Cathy O'Neil "Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy" New York, Crown 2016

[8] Big Brother Watch. "Big Brother Watch Briefing on facial recognition surveillance". June 2020. https://bigbrotherwatch.org.uk/wp-content/uploads/2020/06/Big-Brother-Watch-briefing-on-Facial-recognition-surveillance-June-2020.pdf , consulté, le 07/03/2022

[9] Assemblée Nationale. "Proposition de loi d'expérimentation créant un cadre d'analyse scientifique et une consultation citoyenne sur les dispositifs de reconnaissance faciale par l'intelligence artificielle". https://www.assemblee-nationale.fr/dyn/15/textes/l15b4127_proposition-loi# , consulté, le 07/03/2022

[10] The Perpetual line-up. "Unregulated police face recognition in america". Otober 18, 2016. https://www.perpetuallineup.org/ , consulté, le 07/03/2022

[11] Dr. Joy Buolamwini. "The Algorithmic Justice League : unmasking bias", December 15, 2016. https://medium.com/mit-media-lab/the-algorithmic-justice-league-3cc4131c5148 , consulté, le 07/03/2022

[12] Biometric update.com "Facebook files patent for facial recognition for physical payments". November 13, 2017. https://www.biometricupdate.com/201711/facebook-files-patent-for-facial-recognition-for-physical-payments ,consulté, le 07/03/2022

[13] PYMNTS.com "Facebook Looks at the face in-store payments". November 10, 2017, https://www.pymnts.com/facebook/2017/facebook-patent-facial-recognition-authentication/ , consulté, le 07/03/2020

[14][15] Dr. Joy Bulamwini, Dr. Timnit Gebru "Gender Shades : Intersectional accuracy disparities in commercial gender classification". Sorelle A. Friedler and Christo Wilson. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf , consulté, le 07/03/2020

[16] Statistica. "Meta's (formerly Facebook Inc.) advertising revenue worlwide from 2009 to 2021". https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide/ , consulté, le 07/03/2022

[17] Statistica. "Advertising revenue of Google from 2001 to 2021". https://www.statista.com/statistics/266249/advertising-revenue-of-google/ , consulté, le 07/03/2022

[19] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, Aaron Rieke. "Discrimination through optimization : How Facebook's ad delivery can lead to skewed outcomes". https://arxiv.org/abs/1904.02095 , Last revised 12 Sept 2019, consulté, le 07/03/2022

[20] Lambrecht, A. and Tucker, C. "Algorithmic Bias ? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads". 2018

[22][23] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica "Machine Bias : There's software used across the country to predict future criminals. And it's biased against blacks". May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing , consulté, le 07/03/2020

[24] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner. "How We analysed the COMPAS Recidivism Algorithm". May 23, 2016. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm , consulté, le 07/03/2022

[25] Cédric VILLANI "Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne". Mission parlementaire du 8 Septembre 2017 au 8 Mars 2018. https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf , consulté, le 07/03/2022

[26] Dr. Matt Turek DARPA "Explainable Artificial Intelligence (XAI)". https://www.darpa.mil/program/explainable-artificial-intelligence#_blank , consulté, le 07/03/2022