2021

Les Legs de partialité: Les enjeux éthiques dans l'intelligence artificielle

Christine Ayoh , '21

Les Legs de partialité: les enjeux éthiques dans l'intelligence artificielle

par Christine Ayoh

A senior paper submitted in partial fulfillment of the requirement for the degree of Bachelor of Arts in French and Francophone Studies at Swarthmore College
2021

French and Francophone Studies Section
Micheline Rice-Maximin

Table des Matières

Introduction

Un des champs d'études qui est en train de se développer le plus rapidement c'est l'informatique. Aujourd'hui on peut utiliser l'informatique dans n'importe quel aspect de la vie et c'est la raison pour laquelle beaucoup d'entreprises tentent d'améliorer leurs côtés technologiques soit pour rendre plus facile leur travail soit pour satisfaire ou amuser leur public. Cette amélioration peut se dérouler sous plusieurs formes mais la plupart des améliorations technologiques qu'on voit aujourd'hui sont avec l'intelligence artificielle. L'intelligence artificielle implique la création de modèles pour la machine pour imiter certains processus humains dont la prise de décision, la reconnaissance vocale et la résolution de problèmes. Autrefois, on se préoccupait de la possibilité de créer des machines à l'image des humains, mais à notre époque, on pose une question différente. On sait maintenant qu'on peut faire beaucoup de choses avec l'intelligence artificielle mais où est-ce qu'on fixe les limites ? Cette technologie contient tant de pouvoir et de capacité de changer le monde entier pour le meilleur ou pour le pire.

Un des outils technologiques dans les dernières années un peu controversé grâce à leur immense capacité qui peut être utilisée n'importe comment est l'apprentissage machine. L'apprentissage machine est un sous-champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d'«apprendre» à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune. Il existe deux types d'apprentissage : si les données sont étiquetées, c'est-à-dire que la réponse à la tâche est connue pour ces données, il s'agit d'un apprentissage supervisé sinon il s'agit d'apprentissage non supervisé. Avec la création des outils et algorithmes pour faire des tâches d'apprentissage non supervisés, les machines

gagnent le potentiel d'apprendre les tâches compliquées qu'il n'était pas possible d' apprendre avec les algorithmes précédents. Suivant les progrès, on peut programmer une machine pour apprendre de nombreuses missions dans des domaines complètement différents comme la reconnaissance de la fraude, le regroupement des motifs d'ADN et la construction de systèmes de recommandation. L'apprentissage machine est devenu de plus en plus connu et de plus en plus recherché même s'il y a un danger implicite avec une technologie si avancée. Les craintes n'empêchent pas les gens d'essayer de perfectionner leur technologie dans le but de construire la meilleure imitation des humains, en sachant qu'il n'y a pas assez de connaissances sur le sujet pour pouvoir vraiment les gérer. Les merveilles d'intelligence artificielle ne sont pas toutes connues mais avec beaucoup plus de recherches dans le champ, on découvrira de plus en plus les avantages, les dangers et le savoir-faire pour l'utiliser l'intelligence artificielle correctement.

Avec tous les avertissements contre ces technologies, je m'intéresse toujours à ce champ d'études parce que c' est du jamais vu. Il y aura toujours des nouveautés ou des nouvelles qui ne pouvaient pas être imaginées il y a 50 ans et d'après moi ça c'est magnifique. Cette courbe d'apprentissage qui appartient à ce champ rend le travail toujours intéressant, mais l'autre côté d'une telle courbe d'apprentissage signifie que le champ n'est pas bien régulé ou géré. Les conséquences de travailler dans un champ sous-développé c'est qu'on n'a pas assez de savoir sur les effets du travail et la façon dont la société réagit au travail. Ce fait est absolument vrai pour l'intelligence artificielle et l'apprentissage machine et actuellement, on voit les ramifications de l'utilisation de cette technologie sans les comprendre bien. Ce comportement suscite de grandes discussions sur les enjeux éthiques dans l'intelligence artificielle qui commencent avec la question “Est-ce que c'est éthique d'utiliser cette technologie sans comprendre les ramifications sociales?”. Beaucoup de chercheurs disent

non, et dans cette thèse, je vais examiner les raisons qui soutiennent cette décision. Pour commencer, on peut cadrer les enjeux éthiques dans deux catégories: les problèmes du lieu d'autorité et les problèmes des dilemmes moraux. Le lieu d'autorité concerne la façon dont on utilise la technologie qu'on a construite et qui va la gérer. La concentration est d'assurer que la technologie se trouve “entre de bonnes mains” c'est-à-dire sous la direction d'une entreprise ou un groupe qui a de bons motifs pour le public. Pour une société, c'est difficile de décider ce qui peut constituer être un bon motif et c'est vraiment le problème. L'autre catégorie, les dilemmes moraux, concerne la manière dont on développe cette technologie. On doit examiner qui constitue les équipes des programmateurs et sur quelles données est-ce qu'on base nos algorithmes parce que les préjugés peuvent être reflétés par l'intelligence artificielle, soit les préjugés des membres de l'équipe de programmations, soit les préjugés des données. Ce sont de grandes idées éthiques à analyser dans l'intelligence artificielle et ces idées sont un phénomène mondial.

Dans ce mémoire, je vais explorer le champ d'intelligence artificielle et les problèmes éthiques de lieu d'autorité et des dilemmes moraux qui découlent des avancements dans ce champ. J'utiliserai mes travaux précédents aux LIMSI[1] pour affiner la perspective parce que l'intelligence artificielle est un très grand champ avec de nombreuses applications. Je discuterai des défis éthiques dans le cadre de l'IA spécifiquement le problème de lieu d'autorité et les dilemmes moraux et comment on peut les adresser. Finalement, j'examinerai les préjugés qui sont ancrés dans les fondations des sociétés, principalement les Etats-Unis et la France et un peu le Canada et comment les technologies de l'IA reflètent et révèlent ces préjugés.

[1] Ayoh, Christine. “L'Automatisation d'un robot social .” *Institute for Field Education*, 2019.

Le Problème de lieu d'autorité dans l'IA

Pendant l'automne 2019, j'ai passé le semestre à Paris. Pendant ce semestre, j'ai travaillé dans un laboratoire nommé LIMSI-CNRS. LIMSI, Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, est un laboratoire pluridisciplinaire du CNRS, Centre National de la Recherche Scientifique. Le CNRS est une institution qui finance beaucoup des laboratoires de toutes les disciplines scientifiques et régulent les projets de recherches qui sont produits par ces laboratoires. J'ai travaillé dans le département Communication Homme-Machine sur le groupe Traitement du Langage Parlé (TLP) a LIMSI. TLP se focalise sur le grand thème de traitement de langage naturel, qui implique de concevoir des modèles du langage parlé et de les automatiser. Dans ce groupe, je travaillais sous la direction de la chercheuse et professeure Mme Laurence Devillers et son équipe « Dimension Affectives et Sociales des Interactions Parlées avec des (Ro)bots et Enjeux Éthiques ».

> Les activités autour du thème « Dimension Affectives et Sociales des Interactions Parlées avec des (Ro)bots et Enjeux Éthiques » ont été concentrées sur trois axes : le premier axe porte sur la robustesse de la détection des émotions à partir d'indices paralinguistiques et l'utilisation de ces systèmes dans les interactions avec des robots, le deuxième axe porte sur l'interaction affective avec des machines en utilisant des théories en linguistique sur l'interaction, en sociologie sur les rites sociaux, en psychologie cognitive sur les modèles d'évaluation et la théorie des états mentaux, et enfin le troisième axe porte sur le besoin de réflexions éthiques autour de la modélisation affective et le pouvoir de manipulation par les machines vocales (chatbot, robots sociaux, objets vocaux connectés) dans la société.[2]

Dans cette citation, Devillers démontre la nature interdisciplinaire de ce projet et il en est de même pour la plupart des projets d'IA. Les technologies influences des domaines sociologique,

[2] Devillers, Laurence (2019). "Dimensions affectives et sociales des interactions parlées avec des (ro)bots et enjeux éthiques", *Limsi.fr,* https://www.limsi.fr/fr/recherche/tlp/themes/dimensions-affectives-et-sociales-des-interactions-parlees, 17 novembre 2019.

linguistique et psychologique pour n'en citer que quelques-uns et l'éthique a des implications vraiment réelles dans ces domaines qui influencent alors la société dans son ensemble.

L'équipe a un projet Bad Nudge Bad Robot, c'est dans ce cadre que j'ai effectué mon stage. Bad Nudge Bad Robot est un projet interdisciplinaire avec l'objectif de travailler sur les 'nudges' (suggestions indirectes) dans les interactions parlées entre des robots et des enfants de différents âges et le comportement de ces enfants. Le 'nudge' est un concept de la psychologie comportementale qui traite de la manipulation douce pour inciter les gens à changer de comportement. Cela peut arriver de nombreuses façons mais dans le cas de ce projet, on se focalise sur la manipulation incitative linguistique par les machines affectives. Le but de Bad Nudge Bad Robot est de créer un robot automatisé avec les données issues des expériences susmentionnées en utilisant l'apprentissage machine. Avec une telle expérience, l'apprentissage comporte généralement deux phases. La première consiste à estimer un modèle à partir de données, appelées observations, qui sont fournies à la machine et en nombre fini, lors de la phase de conception du système. L'estimation du modèle consiste à résoudre une tâche pratique, par exemple reconnaître les émotions dans la voix des enfants, la tâche que Bad Nudge, Bad Robot essayait d'automatiser. La seconde phase correspond à la phase de reconnaissance : de nouvelles données, qui n'ont pas servi à l'apprentissage, peuvent alors être soumises au modèle déterminé, afin d'obtenir le résultat correspondant à la tâche souhaitée. Avec l'analyse des données et des interactions, un algorithme serait créé pour développer un robot qui peut interagir et réagir dans n'importe quelle situation. En utilisant ces outils technologiques, l'équipe est optimiste qu'au fur et à mesure le robot obtiendra la capacité de reconnaître quelques émotions ou états mentaux avec la personne à laquelle il parle. Au-delà de cela on utilise les données des expériences pour essayer de trouver les motifs dans le comportement de gens et découvrir qui est le plus

susceptible aux 'nudges' à cause de traits de personnalité qu'on a extraits. Aussi, on peut exploiter les résultats de ces analyses pour tenter de prédire le comportement à l'avenir.

On voit immédiatement que le but du projet Bad Nudge, Bad Robot, est un très grand travail pour les membres de l'équipe, travail qui requiert beaucoup de données et beaucoup de jeunes participants. À chaque étape vers la possibilité d'une création de ce robot, on rencontre des instances ou le projet a la potentialité de dévier et découvrir des informations avec la possibilité d'être nuisible. Même si le risque n'est pas grand, juste la présence de cette possibilité établit le besoin pour un groupe comme le CNRS, qui est chargé d' assumer l'autorité de prendre des décisions morales quand elles surviennent. Si Bad Nudge, Bad Robot réussit à tous ses buts, le groupe pourrait entraîner un robot à reconnaître des profils psychologiques du public en leur parlant et mettre en œuvre la meilleure stratégie pour l'influencer à faire n'importe quoi. De plus, si la stratégie ne marche pas, le robot pourrait la modifier à l'avenir pour obtenir un résultat souhaité. Imaginons que les gens arbitrairement possèdent ce pouvoir ou obtiennent ces données privilégiées et peuvent utiliser cette technologie n'importe comment? C'est une pensée effrayante et c'est une des situations que le CNRS essaie d'éviter. Le CNRS était fondé pour garantir que la technologie interagit avec des gens qui n'ont que des buts bénéfiques pour les êtres humains et aussi acceptés culturellement. Ceci dit, comment est-ce qu'on s'assure que les motifs de cette organisation sont purs? C'est juste un des enjeux éthiques dans l'IA qui s'appelle le problème de lieu d'autorité.

Le problème de lieu d'autorité émerge avec de telles questions. Il y aura peu d' objections avec l'idée qu'il faut des règles en place pour la technologie puissante comme l'IA et ces règles devaient être accessibles et largement acceptées par la société. Une tâche plus contestée c'est le processus de trouver une autorité qui est très fiable pour prendre des décisions importantes

comme les règles qui seront nécessaires pour réguler le champ croissant d'IA et comment ces règles seront imposées. Ce pouvoir est très important car c'est trop facile avec ces technologies de commettre des fautes fatales. Le pouvoir de faciliter ou de gêner ces fautes reste avec le lieu d'autorité et c'est pourquoi ce problème éthique est insurmontable et existe sur une échelle nationale et aussi sur des échelles plus petites comme une échelle institutionnelle. En connaissant les risques inhérents dans ce domaine, en 2018 la France a annoncé son plan "pour faire de la France un pays leader de l'intelligence artificielle"[3]. Cette proposition rappelle aux chercheurs, ingénieurs et développeurs de garder les éthiques en tête dans leur travail pour encourager la transparence entre eux et le public pour l'inviter à faire partie des décisions de la façon d'utiliser mieux l'IA pour la société. L'importance de cette transparence est démontrée par l'exigence d'un "développement [de] procédures, outils et méthodes permettant d'auditer ces systèmes afin d'en évaluer la conformité à notre cadre juridique et éthique"[4]. Ces sentiments sont nécessaires mais ils tardent à devenir un problème de magnitude nationale. Avant cela, le CNRS s'est focalisé depuis 1994 dans la création de COMETS. COMETS qui est "le Comité d'éthique du CNRS est une instance consultative indépendante, placée auprès du conseil d'administration du CNRS [qui] a pour missions de développer la réflexion sur les aspects éthiques suscités par la pratique de la recherche, de formuler des recommandations et de sensibiliser les personnels à l'importance de l'éthique"[5]. Étant une instance consultative indépendante, le CNRS assure qu'il n'y a pas de conflits d'intérêts à analyser des implications éthiques dans les projets des laboratoires donc veut garder à cœur l'intérêt de la société. On ne sait pas vraiment qui constitue COMETS mais le dirigeant Jean-Gabriel Ganascia est bien éduqué en IA et appartient aux autres organisations

---

[3] Villani Cédric. Éditeur Inconnu, *Donner un sens à l'intelligence artificielle: pour une stratégie Nationale européenne: Mission Parlementaire du 8 septembre 2017 au 8 mars 2018*, 2018.

[4] IBID

[5] "Jean-Gabriel Ganascia." *CNRS*, www.cnrs.fr/en/person/jean-gabriel-ganascia.

d'éthique dans la technologie. Soutenu par ses certifications, c'est plus facile de lui faire confiance ainsi qu'à la vision éthique de son équipe. L'équipe de COMETS transmet sa sagesse éthique avec des guides et des conseils sur la manière de pratiquer une recherche intègre et responsable mettant les labos CNRS entre de bonnes mains. Avec le soutien des organisations comme le CNRS et la nouvelle mission de la France, l'éthique en IA est un concept très répandu et respecté aujourd'hui bien qu'il y ait une dizaine d'années que ces organisations se soient créées pour régler les problèmes éthiques évidents et on parle toujours des mêmes enjeux. Malgré la présence des organisations comme celles qui gèrent Bad Nudge, Bad Robot, cette structure n'est pas toujours la même d'un groupe à l'autre.

Sans les décrets et les régulations d'un corps consultatif comme le CNRS, beaucoup de projets de recherches scientifiques croisent des zones de flou. On combat cette difficulté en maintenant la transparence dans le travail et en évitant des conflits d'intérêts dans les groupes de programmation et les corps consultatifs des instituts de recherches. Ce style de gouvernance n'est pas la seule méthode de gestion des éthiques dans l'IA, mais selon moi c'est la meilleure. Qu'est-ce qui se passe dans un conglomérat qui n'établit pas de règles éthiques claires? Depuis sa conception, Google était à la première ligne concernant les avancées technologiques. Tout en étant une entreprise publique, cela a consolidé le statut de Google en tant que méga puissance pas seulement dans la technologie mais aussi dans la politique étrangère. Nos vies actuelles ne fonctionnent pas sans les initiatives de Google et c'est la raison pour laquelle on permet à Google d' accéder à tout; même aux parties les plus intimes de nos vies. Avec le prestige qui accompagne le nom de Google vient une puissance technologique immense qui, mal utilisée, a la capacité de ruiner des vies. Sans appartenir à l'entreprise, c'est presque impossible de savoir comment vraiment Google gère tous les avancements pour lesquels elle est responsable. On ne sait pas

quels sont les critères qu'un projet doit posséder pour devenir autorisés pour un usage plus répandu et quelle est la définition distincte d'un travail éthique. Le manque de transparenc de la part de Google nous laisse dans le noir et c'est impossible de quantifier le problème de lieu d'autorité dans ce cas. En comparant Google avec le CNRS, on voit qu'on ne sait pas les règles pour les travaux avec IA, on ne sait pas qui porte la responsabilité d'assurer que ces règles sont respectées, et on ne sait pas du tout ce que Google voit comme éthique ou pas. Ce manque de connaissances provoque une zone de flou tellement dangereux car cela affecte tout le monde dans n'importe quel coin et peut mettre beaucoup de gens, spécifiquement les minorités, en péril.

En décembre 2020, Google a licencié Timnit Gebru, une des dirigeantes de l'équipe d'IA Responsable. Gebru était une chercheuse très respectée dans le monde de la recherche sur l'éthique de l'IA et elle soulignait la connaissance des dangers des tâches communes dans l'IA comme la reconnaissance faciale et la modélisation du langage. C'était ses écrits révolutionnaires[6] qui traitaient des risques de la modélisation du langage qui ont poussé Google à la licencier. Un élément clé de l'activité de Google est les modélisations du langage[7] alors ses écrits qui parlaient des risques interfèrent directement avec les affaires de Google et subséquemment interfèrent avec leur finance et leur statut. Dans un entretien avec France 24, Laurence Devillers parle de cet événement en disant “qu'il ne faut pas penser que plus de données veut dire un système plus intelligent… Pourquoi faire confiance à ces modèles de langage nourris non pas par des textes sélectionnés, mais par les données d'Internet représentant beaucoup de fake news ?”[8]. Elle partage la même perspective que Gebru qui met en avant les

---

[6] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3442188.3445922

[7] Hao, Karen. “We Read the Paper That Forced Timnit Gebru out of Google. Here's What It Says.” *MIT Technology Review*, MIT Technology Review, 7 Dec. 2020

[8] SEIBT, Sébastian. “Timnit Gebru, licenciée par Google pour sa vision 'éthique' de l'intelligence artificielle ?” *France 24*, France 24, 8 Déc. 2020

faits que les modélisations que Google utilisent sont absolument biaisées et la philosophie de Google du 'plus, c'est mieux' effectivement ne profite qu'aux géants du Net comme le club des GAFA (Google, Apple, Facebook, Amazon). C'est impossible d'ignorer “les risques de cette course aux bases de données toujours plus grandes”[9] alors les actions de Google nous montrent ses priorités. Ce récit démontre le besoin d'un corps consultatif pour au moins parler des implications éthiques de ces modélisations et trouver un équilibre entre le désir d'une entreprise de faire grandir des données et les conséquences potentielles pour la société. C'est contradictoire et ironique que Google ait créé une équipe pour faire des recherches sur l'éthique en IA, mais quand cette équipe fait le travail demandé, Google prend des mesures pour gêner des efforts pour introduire des pratiques éthiques en renvoyant de ses entreprises une dirigeante très respectueuse de l'éthique. Clairement, Google n'est pas fiable comme un lieu d'autorité et promeut ses intérêts égoïstes par rapport à ceux de la société en général. Non seulement cette histoire de Google montre l'importance d'avoir un lieu d'autorité responsable et transparent, mais elle introduit un autre enjeu éthique dans l'IA nommé les dilemmes moraux dans l'IA.

Les (nombreux) Dilemmes moraux dans l'IA

Dans le domaine de la philosophie, les problèmes éthiques que j'appelle les dilemmes moraux sont divisés en deux, soit les dilemmes éthiques, soit les problèmes de désarroi moral. Les dilemmes éthiques se produisent de trois façons: quand il y a des problèmes dont on ne sait pas la réponse correcte, quand il existe au moins deux réponses correctes possibles et quand l'action de réaliser une de ces réponses correctes inhibe l'autre. En revanche, les problèmes de désarroi moral émergent quand on sait la réponse correcte mais il y a aussi des inhibitions qui

[9] IBID

nous empêchent de la choisir. En groupant ces enjeux dans une catégorie “dilemmes moraux”, on peut analyser les ennuis de l'IA plus généralement. Ces dilemmes moraux impliquent les organisations qui produisent cette technologie, les équipes qui travaillent sur ces projets sensibles et l'industrie de l'informatique en général. Pourquoi est-ce que les dilemmes moraux sont applicables au domaine de l'intelligence artificielle? Le but de l'intelligence artificielle dès qu'il est créé est de former des machines à l'image humaine. Les machines pourraient penser, comprendre et réagir en imitant les processus cognitifs humains. Le problème d'essayer de faire des machines comme des humains est que ce n'est pas possible. Malgré cela, les chercheurs et programmeurs essaient de rendre les machines le plus proche possible des humains. Cela mène les mêmes chercheurs et développeurs à programmer leur biais implicite dans les machines. Il n'est pas possible de programmer une machine pour oublier des biais comme on tente de le faire avec des humains. Il n'existe pas de processus pour désapprendre ces biais ou les stéréotypes qu'on a donnés à la machine. Ce fait est la raison pour laquelle les dilemmes moraux dans l'IA existent et on doit faire tout ce qu'on peut pour les combattre. En créant des machines à préjugés, on met la vie de tous les gens qui appartiennent aux communautés les plus vulnérables en grave danger.

Pour mieux comprendre la gravité d'une machine à apprendre des préjugés, on doit vraiment comprendre le potentiel des machines “intelligentes” et ce qu'est l'intelligence artificielle et comment elle est utilisée dans des machines. La Commission Nationale Informatique & Libertés (CNIL) reconnaît l'importance de définir et interpréter les mots d'IA et les processus ci-dessus pour tenir le public au courant et l'informer des enjeux éthiques dans les algorithmes d'IA. Pour mettre la compréhension des algorithmes d'IA en perspective, le CNIL a fait une enquête en 2017 pour découvrir le sentiment général d'IA et ses algorithmes en France et

elle a conclu “si 83% des Français ont déjà entendu parler des algorithmes, ils sont plus de la moitié à ne pas savoir précisément de quoi il s'agit (52%)”[10]. Cette statistique est moins que satisfaisante et dans son rapport *Comment permettre à l'homme de garder la main*, le CNIL explore comment “parvenir à une base de discussion minimale et opératoire qui permette de tracer pragmatiquement le périmètre des algorithmes et de l'intelligence artificielle sources de questions éthiques et de société cruciales”[11]. Le rapport divulgue les bases d'IA et de l'apprentissage machine, comment on utilise les algorithmes dans ces applications et pourquoi il faut être prudent avec ces technologies afin de créer une base de compréhension pour le public.

En analysant les principes fondamentaux de ces technologies, la prochaine étape logique est de parler des algorithmes qui soutiennent ces technologies. “Algorithme” est un terme général pour parler de la méthode pour résoudre un problème. En IA le problème qu'on veut résoudre c'est la façon de communiquer les buts à la machine. L'algorithme saisit des données et fabrique le résultat qu'on cherche. Chaque algorithme est différent et peut se réaliser dans des mises en œuvre diverses. Dans ce cas, les développeurs qui ont accès au code sont les seuls qui peuvent contrôler la manière dont la machine interagit avec les données. Cela donne aux développeurs beaucoup de pouvoir et introduit la question de préjugés dans les algorithmes. Ci-dessus on a discuté du caractère inévitable d'avoir les préjugés dans les systèmes créés par l'homme et maintenant on en examinera les raisons. Quand les développeurs reçoivent un projet, on leur donne tous les objectifs des projets des organisations de recherches pour qu'ils travaillent. Même s'ils ont des tâches particulières à accomplir, tous les aspects de la réalisation sont à leur discrétion. Ils doivent trouver une méthode de l'implémentation pour atteindre les objectifs et c'est possible d'avoir du mal à concevoir des algorithmes. Peut-être que les développeurs sont

[10] Victor Demiaux. “Comment permettre à l'homme de garder la main ?” CNIL, 2017.
[11] IBID

inhibés par les buts d'organisation de recherche ou peut-être que les lois ou politiques dans la région empêchent leur implémentation. Tout ce qui gêne les développeurs, le problème de savoir comment résoudre un problème mais pas comment l'implémenter, est un problème de désarroi. Les problèmes de désarroi sont une façon d'impliquer l'éthique dans les algorithmes des développeurs; l'autre c'est que le sort d'un logiciel reste entièrement aux mains d'un développeur ou son équipe. Ce contrôle sur le logiciel signifie que la façon de penser des développeurs sera effectivement reflétée dans le logiciel fini. Les forces extérieures n'ont pas beaucoup d'influence sur cette circonstance car c'est leur mode de vie. Des logiciels nous montrent beaucoup de la personnalité des développeurs qui les ont créés et dans plusieurs cas, cela inclut les biais implicites. Le biais implicite est inévitable et non intentionnel; on ne peut pas le gérer. Ces biais existent partout et comme humains, on sait comment s'adapter aux biais et ne pas répondre aux biais nuisibles. La machine n'a pas de processus cognitifs comme nous et ne sait pas quand les biais sont introduits dans son code. Elle ne peut que réagir à la façon dont elle est programmée, et si le code inclut des biais, la machine sera biaisée. Le biais implicite est un phénomène auquel on ne peut pas échapper et si la société veut vraiment avancer la technologie d'IA, il faut trouver une manière de gérer les biais implicites dans les algorithmes.

Il existe d'autres types de biais qui sont applicables à l'IA. Contrairement au biais implicite, le biais explicite est intentionnel. Il reflète nos sentiments et nos croyances et c'est un choix conscient d'agir sur ces croyances. Le biais explicite est plus tabou que le biais implicite car c'est plus facile de l'identifier dans le monde général et aussi dans l'IA. Malgré le fait que ce type de biais est largement évité, il joue toujours un rôle dans les projets d'IA. Les organisations de recherche, en décidant ce qu'elles veulent rechercher, doivent aussi considérer quelles sortes d'algorithmes elles veulent développer et à quelles questions elles veulent répondre. Ces

considérations ont des biais inhérents. Les laboratoires de recherche choisissent des projets ou des questions qui les intéressent. Ce qui les intéresse est modelé par leurs milieux, leurs expériences et leurs positions dans la société. Bien qu'il soit possible de combattre les biais que nos vies nous ont donnés, c'est très difficile. Même si les chercheurs dans ces laboratoires n'ont pas l'intention d'introduire des facteurs de confusion dans leur travail, leurs biais explicites sont présents de toute façon dans les projets et les initiatives qu'ils suivent. La prochaine étape après que les questions de recherches soient choisies c'est que les équipes de recherche doivent réfléchir aux réponses éthiques de ces problèmes. C'est difficile d'analyser toutes les parties éthiques en se concentrant sur un problème et cela dévoile les dilemmes éthiques dans l'IA. Ils doivent tenter de savoir s'il y a plusieurs réponses à ce problème ou juste une. S'il y en a juste une, comment est-ce qu'on s'assure que les développeurs produisent la solution de la meilleure manière? S'il y a plusieurs réponses, quelle réponse est-ce qu'ils choisissent? Comment élaborer les critères pour comparer les effets futurs de ces réponses et choisir ceux qui sont avantageux? Peut-être que les réponses sont également avantageuses ou les avantages d'une réponse inhibent les avantages d'une autre. Puis qu'est-ce qu'on fait? C'est presque impossible de trouver une bonne solution à ces questions éthiques et encore on se retrouve dans une zone de flou.

En dehors des organisations de recherches et des développeurs qui travaillent pour eux, l'industrie d'informatique en gros et spécifiquement les constituants qui se focalisent sur l'IA contribuent à cette zone de flou. L'industrie exhibe le biais systémique et selon moi, c'est le biais le plus dangereux. Le biais systémique consiste en des préjugés contre des groupes minoritaires. Ces préjugés sont reflétés dans chaque partie de la société. En ce qui concerne l'IA, le biais systémique est présent dans les données qu'on utilise dans les algorithmes. On a déjà discuté des algorithmes que les développeurs créent mais pour que l'algorithme puisse rendre des résultats on

doit lui fournir des données. Ou est-ce qu'on trouve ces données? La réponse est partout. Dans le monde actuel, il y a plusieurs options pour trouver des données pour n'importe quel projet qu'on doit suivre. Les organisations de Big Data sont les plus grands fournisseurs de données et algorithmes qui fonctionnent avec ces données. Il existe d'autres sources de données et d'autres créatures d'algorithmes, mais l'omniprésence de Big Data rend ces méthodes et données les plus connues. On a examiné le problème de transparence avec une entreprise de Big Data comme Google. Ce manque de transparence est un grand contributeur au biais systémique dans les données mais il y a rien a faire pour le public pour combattre ce biais. Les organisations de Big Data peuvent gérer des données préjugées et malveillantes ou créer des algorithmes sans transparence afin qu'on ne voit pas le fonctionnement interne ou le processus de pensée des développeurs. Ce comportement clandestin des entreprises Big Data nous conduit soit à leur faire confiance aveuglément, soit à se méfier de tous les produits de ces entreprises. D'après moi, l'option la plus consciente c'est de se méfier de ces entreprises.

Pour mettre en perspective, le pouvoir destructif de Big Data, on peut réexaminer Google. Google est un grand partisan de la reconnaissance faciale depuis qu'il a lancé son logiciel en 2015. Même à cette époque, il avait déjà rencontré des problèmes avec cette technologie, mais il a insisté pour laisser la technologie pour la consommation de masse. Google Photos a été mise à jour avec la capacité de reconnaître des objets dans les photos et de les catégoriser pour le public. Un ingénieur logiciel a consulté son album de Google Photos peu après cette mise à jour seulement pour trouver qu'un album avec lui et ses amis étaient catégorisés comme des gorilles. On doit garder en tête que l'ingénieur du logiciel et ses amis dans cet album sont noirs. On peut spéculer sur la cause de cette faute répugnante mais la seule réponse correcte c'est la présence de biais dans les données que Google a fourni à ces algorithmes. Les développeurs de Google ont

utilisé l'apprentissage machine pour trouver des caractéristiques similaires entre les photos qui ont nourri l'algorithme. Les données ont reflété le stéréotype que les noirs ressemblent à des singes et potentiellement ont inclus des images racistes des caricatures de noirs; on ne sait pas maintenant et on ne saura jamais. Ce qu'on sait c'est que cet ingénieur avait eu une expérience raciste avec la technologie de Google qui est répandue et utilisée comme infaillible. Le "problème" était réglé peu après cette affaire mais il aurait pu y avoir des conséquences très réelles et très dangereuses pour chaque personne noire dans le monde, si la technologie avait été utilisée d'une autre façon. Il existe beaucoup de raisons pour lesquelles cette erreur technologique est problématique, mais aux mains d'un conglomérat comme Google, c'est embarrassant et méprisable que leur logiciel reflète des préjugés et des stéréotypes aussi dégoûtants. Ces biais et ces enjeux éthiques existent dans l'IA n'importe où les algorithmes et les données sont développées ou utilisées et on a exploré la gestion de ces problèmes dans des organisations différentes. Un grand déterminant du style de gestion dépend de la société où la technologie est utilisée. Même si ces enjeux restent les mêmes, ils sont réglés très différemment. Les différences entre le style de gestion de la France et celui des Etats-Unis représentent une divergence des idées et normes sociétales si fortes que la France le voit comme une menace.

Les Idées américaines menacent-elles la cohésion française?[12]

C'est évident que les algorithmes et les systèmes d'IA reflètent les biais qui prévalent dans les sociétés et quelques fois révèlent des préjugés dans les données qu'on ne connaît pas. En sachant cela, comment ces préjugés et stéréotypes sont-ils devenus si ancrés dans la société? La réponse est que les développements historiques qui façonnent les sociétés sont quelques fois

[12] Norimitsu Onishi. "Les Idées américaines menacent-elles la cohésion française?" *New York Times*, 9 Feb. 2021

subjectifs et ces subjectivités deviennent des biais qui déforment les données qui soutiennent la façon dont l'IA comprend le monde. C'est impossible d'analyser et de discuter tous les composants qui ont contribué à ces subjectivités, mais je mettrai l'accent sur les éléments les plus pertinents qui créeront les motifs des discriminations pour tenter de comprendre la science derrière les biais. En examinant la France et les États-Unis, on voit les héritages de ces pays qui ont ouvert la voie pour que ces manières de pensées biaisées soient normales. En France, les principes de liberté, égalité, fraternité, et laïcité contribuent à ces manières de pensées. Ces principes étaient enracinés dans la fondation de la France. L'expression «liberté, égalité, fraternité» est le slogan officiel de la France d'aujourd'hui et ses racines se trouvent dans la Révolution française dont les événements ont engendré la première constitution de la France. En 1789, elle a établi la Déclaration des droits de l'homme et du citoyen pour redonner les droits que la monarchie avait volé au peuple. Il y avait une crainte de retourner à cette époque de répression et ces libertés sont un type d'assurance que la société n'y retournera jamais. La déclaration promet que chaque personne dans la société française est libre et égale et elle révèle les spécificités de ces principes de liberté et égalité. L'idée de fraternité se présente quand les français ont établi les principes de la liberté et de l' égalité. Car tout le monde mérite les mêmes droits, peu importe l'origine. Avec cette devise, la France a institué la croyance selon laquelle être français signifie l'accès inconditionnel et instantané à une communauté où tout le monde est égal et libre. Alors que le pays a considérablement changé depuis la création de cette devise, cette phrase reste un symbole politique important et reste vraiment pertinente dans la société d'aujourd'hui. En revanche, les États-Unis ont été fondés sur la devise de «life, liberty, and the pursuit of happiness» (la vie, la liberté, et la poursuite du bonheur) en plus du principe de la liberté d'expression. Cette devise est inscrite dans la Déclaration de l'Indépendance des

États-Unis et a été constituée dans une période tournante pour le pays, après sa révolution contre le Royaume-Uni. Elle est aussi considérée comme inspiration pour la constitution[13]. La Déclaration dit “We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable rights, that among these are Life, Liberty and the Pursuit of happiness”[14], phrases qui marquent le début de discussion des droits des citoyens des États-Unis et la devise contient les trois droits les plus importants qui définiront la société. Ces périodes de débats sur les droits civiques sont significatives et précisent toujours les principes importants et durables pour les sociétés. On voit immédiatement l'accent sur les droits innés de l'homme comme principes fondateurs mais bien que les motivations qui soutiennent ces principes soient les mêmes, elles existaient très différemment dans ce pays. Les principes fondateurs de ces pays sont parties intégrantes de la formation de la société qui, en conséquence, influencent les normes sociétales et la perception des groupes différents. Si on regarde les motifs entre les groupes qui historiquement étaient affectés par cette perception, on voit l'émergence de groupes minorisés.

Ci-dessus, on a discuté des droits innés de l'homme que la France et les États-Unis ont donnés à leurs citoyens. Bien que les gouvernements croient généralement en les mêmes droits, la différence entre ces pays est que la France croit que les droits sont innés, chaque personne est née avec ces droits et personne ne peut les lui retirer. En revanche, les États-Unis, eux, croient que ces droits sont inaliénables, que Dieu les a donnés à chaque personne. Cette différence est la plus grande entre ces pays, et à mon avis, elle démontre l'importance du christianisme dès son origine pour les États-Unis et de la laïcité pour la France. En théorie, la France a rejeté l'idée de

---

[13] “Life, Liberty and the Pursuit of Happiness.” *Wikipedia*, Wikimedia Foundation, 19 Apr. 2021, https://en.wikipedia.org/wiki/Life,_Liberty_and_the_pursuit_of_Happiness

[14] “The Declaration of Independence.” *National Archives and Records Administration*, National Archives and Records Administration, www.archives.gov/founding-docs/declaration

privilégier officiellement une religion sur une autre, alors la laïcité en France a été également introduite dans la loi de 1905 qui a ordonné la séparation entre l'État et l'Église. Cette loi était une addition à l'article dix de la Déclaration française et reconnaissait toutes les religions comme égales. Ce principe a évolué depuis sa création et il est devenu si important pour la société française à tel point que le gouvernement l'a inclus dans la constitution de la cinquième république et la laïcité était décrite comme une de ses valeurs. Par contre, la séparation entre l'État et l'Église est sous-entendue dans la constitution des États-Unis, mais en réalité, ce principe n'est pas vraiment respecté comme on l'a déjà vu. Au lieu de la laïcité comme un principe élémentaire pour la société, les américains se concentrent sur une de ces libertés fournie par leur Constitution: la liberté d'expression. En France, ce pouvoir n'est pas forcément garanti à cause du principe de laïcité. En réalité, la laïcité qui était introduite pour établir le sécularisme en France finit par privilégier les français blancs et chrétiens. On verra comment les principes de christianisme et de laïcité serviront à étendre les empires respectifs de ces deux nations. En prenant des chemins différents, les deux pays arrivent aux mêmes conclusions: la préférence pour les blancs dans leurs sociétés.

Ainsi au XIXe siècle, les français et les américains en sécurité dans leurs communautés ont engagé leurs stratégies impérialistes basées sur leurs racines historiques pour étendre leurs empires; les stratégies sont différentes mais les motifs existent toujours. La France sous l'influence des principes de liberté, égalité et fraternité s'est embarquée dans des missions civilisatrices pour que les “sauvages” puissent devenir civilisés et prendre part aux valeurs, à la culture, et la langue française. Les français veulent établir la supériorité de leur culture d’«être français» chez les indigènes. Sans l'aide de la France, ces populations ne méritaient pas les libertés accordées aux “français français” d'où la mission civilisatrice de la France. Malgré le

désir de transformer ces gens en leur image, les français veulent les distinguer des vrais français, qui méritent le droit de fraternité, des français assimilables en ne donnant pas à ceux et celles ancien.nes colonisé.es la citoyenneté française. C'était une façon de les rejeter en leur imposant leur style de vie. Ce comportement des français a créé une vraie distinction entre les deux groupes et provoque la question de ce qu'il signifie d'être un citoyen ou même plus une personne. Pour les français, la citoyenneté est un droit que chaque personne mérite dès sa naissance; refuser à quelqu'un le droit de citoyenneté c'est révoquer leur statut de personne et nier leurs droits innés. Similairement à la France, les États-Unis désiraient imposer leur valeurs à la population autochtone de “leur” territoire. Selon la doctrine de «American exceptionalism»[15], qui dit que les libertés qui sont implantées dans la société américaine à cette époque confirment leur totale supériorité sans aucune exception. Les américains voyageaient à l'Ouest en apportant leurs idées et façons de vivre “supérieures” en croyant que ce caractère exceptionnel est la raison pour laquelle c'était leur destin “tombé du ciel” d'étendre leur empire ou en d'autres mots c'était leur “manifest destiny”. Comme toujours ils recourent au christianisme comme excuse et explication de leur comportement. En étendant leur empire terrestre, les américains ont volé la terre des premiers américains, ils les ont forcé à assimiler leur culture supérieure, et au fur et à mesure, ils ont commis un génocide au nom de Dieu. Effectivement, les américains n'ont pas du tout le respect pour les droits inaliénables des autochtones. Comme les français, les américains considèrent ces droits comme donnés à chaque personne. En fait, les Américains vont plus loin en disant que ces droits, comme leur destinée manifeste, sont donnés par Dieu. Évidemment les autochtones, n'ont pas accès à ces droits ou bien ils auraient été respectés. La seule conclusion à

[15] Ian Tyrrell. “What, Exactly, Is 'American Exceptionalism'?” *The Week - All You Need to Know about Everything That Matters*, The Week, 21 Oct. 2016, https://theweek.com/articles/654508/what-exactly-american-exceptionalism

laquelle on puisse arriver est qu' à travers les yeux des Américains, les autochtones ne méritent pas de droits inaliénables car ils ne peuvent pas être considérés comme des personnes.

En continuant à travers la lentille de l'édification de l'empire, on passe à une époque plus récente. Le but de nier à des gens leurs droits humains est le même, mais les deux empires l'accomplissent différemment. Aux États-Unis, après le genocide des autochtones, les américains avaient besoin d'un autre groupe à subordonner. Cette fois, cette subordination est justifiée par des buts capitalistes eux-mêmes soutenus par le christianisme. Pour l'enrichissement de l'économie de l'empire américain, ils avaient besoin d'une nouvelle main d'œuvre mais pas n'importe laquelle, ils avaient besoin d'une qu'ils pourraient priver de ses droits humains. Ils ont choisi de participer au commerce des esclaves pour obtenir la main d'oeuvre la moins chère possible pour que l'empire puisse profiter le plus possible de ce groupe de gens. Automatiquement, ces gens mis en esclavage ne sont pas considérés comme des personnes comme les américains, mais ils sont considérés comme une propriété et effectivement, on ne peut pas donner de droits humains à une propriété. Il semble controversé d'être un pays fondé sur le christianisme qui traite également les créations de Dieu comme la propriété. Dans son livre Stamped from the Beginning, Ibram X. Kendi écrase la contreverse en expliquant qu'il y avait un influx des idées théologiques racistes qui étaient importantes pour justifier les nombreuses violations des droits de l'homme dans la pratique de l'esclavage et qui sont intrinsèquement contraires à la doctrine chrétienne. Il y avait des pasteurs qui ont promu l'esclavage comme un processus nécessaire pour sauver les âmes sombres des Africains afin qu'ils deviennent blancs. Bien que la subordination des gens mis en esclavage ait été officiellement interdite en 1863, les attitudes contre ces gens créées pendant la période de l'esclavage durent encore dans la societé américaine à cause de la racialisation et du blanchiment des Européens, liberté, civilisation,

rationalité et beauté (“racializing and whitening Europeans, freedom, civilization, rationality, and beauty”)[16]. Entre-temps, la France a aussi engagé plus de tactiques pour agrandir l'empire avec des objectifs plus politiques. Au XXe siècle, en accord avec l'Allemagne, la France de Vichy a participé à l'extermination des juifs européens pendant la Shoah. La raison derrière cette extermination est simple, Hitler et ses disciples obsédés par l'idée d'une race pure allemande, la race Aryenne, ont décidé que les gens qui ne montrent pas ces caractéristiques de pureté ne méritent ni les droits humains ni la vie.[17] La France comme disciple permettait et aidait l'Allemagne à commettre ce genocide car elle croyait aussi que seule la race Aryenne méritait le statut de personne. En dépit de la concentration de la Shoah sur les juifs, il n'est pas improbable de dire que les pays qui soutiennent cette idéologie croient que n'importe quelle personne qui n'est pas blanche ne mérite pas le statut de personne. Toutes les personnes à qui on a nié ces droits appartiennent aux groupes minorisés et maintenant, ces groupes sont les mêmes qui souffrent des biais de machine. A toute époque d'extension de l'empire, les violations des droits sont prédites, en fait pour construire ces sphères d'influence, les autres doivent être subordonnées. A chaque époque, des outils sont développés pour améliorer la construction de l’empire et de cette manière, on voit l'intersection entre le développement de la science et de la technologie et le développement de l'empire. La science a toujours été la partenaire de la conquête et de l'élargissement de l'empire. En révoquant les droits humains de certains groupes, la France et les États-Unis réalisent seulement les buts de l'édification de leurs empires. Comme la conquête et l'édification de l'empire modifiait les paysages réels, l'IA change le paysage de ce que signifie une science définitive.

---

[16] Ibram X. Kendi. *Stamped from the Beginning: the Definitive History of Racist Ideas in America*, Bold Type Books, 2017, pp. 1–11.

[17] History.com Editors. *The Holocaust*. 14 Oct. 2009, www.history.com/topics/world-war-ii/the-holocaust

Dans le rapport *Comment permettre à l'homme de garder la main?* par le CNIL en 2017, l'IA d'aujourd'hui est décrite comme "un sujet de controverses plus ou moins explicites entre chercheurs en intelligence artificielle, entrepreneurs et prescripteurs d'opinions diverses dans le domaine des technologies."[18] Les chercheurs d'IA et les professionnels de la technologie sont en combat direct avec les entrepreneurs qui poussent leurs idées et visions pas informées pour que leurs entreprises puissent profiter économiquement. Dans un monde de plus en plus mené par ses propres intérêts, ce comportement est normal mais avec l'ignorance des entrepreneurs qui traitent l'IA comme une science définitive, la création des technologies nuisibles deviendront de plus en plus communes. En traitant l'IA comme une science définitive, on méprise des normes déformées et des attitudes biaisées ancrées dans la société que l'IA peut apprendre et répliquer. Ce comportement étend la sphère d'influence des groupes historiquement privilégiés, en ce cas les hommes blancs. C'est important de noter que les dirigeants des entreprises intéressées dans les techniques d'IA aussi sont des hommes blancs. Aux États-Unis, plusieurs études et projets sont consacrés à trouver et régler les biais dans les méthodes différentes de l'IA. En reconnaissant leur histoire méprisable, les chercheurs américains visent à retirer les biais et préjugés contre les groupes minorisés reflétés dans les algorithmes et les données. Cette conscience permet au domaine américain de l'IA de mieux comprendre le problème et de développer des moyens de le gérer qui ne nuisent pas aux gens. En revanche, la France en reconnaissant son histoire méprisable a décidé de ne pas s'attarder sur les résultats de ces événements ou assurer que les biais systémiques instillés par l'histoire du pays seraient discutés et compris. Elle a décidé de supprimer les données qui montrent ces biais systémiques après la Shoah et de déclarer que tout le monde est égal sur l'étiquette «français» sans définir ce que signifie cette égalité. Dans son

[18] Victor Demiaux. "Comment permettre à l'homme de garder la main ?" CNIL, 2017, www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf

article, dont j'emprunte le titre pour ce chapitre, Norimitsu Onishi parle de la circulation d'idée que les théories américaines “notamment sur la race, le genre, le post-colonialisme”[19] sont une menace. Il dit “la menace serait existentielle. Elle nourrit le séparatisme. Ronge l'unité nationale. Encourage l'islamisme. Porte atteinte au patrimoine intellectuel et culturel de la France”[20]. C'est évident que l'unité nationale de la France dont Onishi parle ne considère pas les gens blessés par ce patrimoine. La France était structurée pour privilégier un groupe seulement et est en train de défendre son histoire en l'appelant patrimoine. Il apparaît que ce soit un crime pour la France de se développer et de cultiver son héritage pour vraiment refléter que c'est un pays fondé sur les principes de «liberté, égalité, fraternité». Au lieu de cela, elle redouble sur ses principes rabattus et ignore les motifs qui sont en train de diviser en deux sa société. Les États-Unis et la France répondent aux problèmes systémiques qui se produisent à cause de leur histoires de manières opposées mais en gros, ces manières génèrent les mêmes résultats; biais systémiques contre les groupes minorisés qu'ils ont historiquement assujettis. Onishi démontre que c’est l'attitude de beaucoup de français.

> L'État français ne recueille pas de statistiques raciales, illégales, dans le cadre de son engagement affiché en faveur de l'universalisme et du traitement égal de tous les citoyens au regard de la loi. Mais pour nombre de spécialistes de la question raciale, cette réticence s'inscrit dans une longue histoire de négation du racisme en France, du passé colonial et de la traite négrière du pays.[21]

Selon eux, la discussion de la race n'est pas nécessaire grâce à la croyance que chacun a une identité commune comme français mais en éliminant la race on ne peut pas voir les connexions raciales et les motifs des préjugés qui sont représentatifs des comportements et convictions du passé. Les États-Unis prennent le chemin opposé et grâce à cette décision, on a la capacité de

---

[19] Norimitsu Onishi. “Les Idées américaines menacent-elles la cohésion française?" *New York Times*, 9 Feb. 2021

[20] IBID

[21] IBID

parler de ce biais et des causes de biais et on peut essayer d' enlever les subjectivités dans les données qui produisent les biais.

Le biais dans l'IA est discuté largement aux États-Unis et est considéré comme un sujet distinct de recherche dans le domaine. En plus des groupes minorisés qu'on a examinés, les noirs et les amérindiens aux États-Unis, les biais existent contre à peu près n'importe quel groupe minoritaire. Dans son livre Race After Technology, Ruha Benjamin discute des questions de l'iniquité artificielle et se demande si les robots peuvent être racistes. Elle donne un exemple des chercheurs d'une organisation de l'Hong Kong et de l'Australie qui s'appelle Youth Laboratories. Ils ont décidé de créer la première compétition de beauté jugée par les robots dans un effort de trouver un résultat objectif. Les résultats ne sont pas surprenants; toutes les finalistes de la compétition sauf une avaient la peau claire. Les chercheurs étaient consternés mais c'était évident que les robots ont appris les biais de quelques développeurs de l'équipe et ils ont aussi appris la préférence générale de gens avec la peau claire que toutes les données leur ont fourni. Le but de cette technologie de Youth Laboratories était d'apprendre l'information de la santé des gens, de leurs photos et aussi d'identifier les problèmes de couleur de peau pour que le laboratoire puisse recommander une façon de ralentir le processus de vieillissement et d’améliorer la santé. Bien que les objectifs de ce projet soient bons et pourraient être utiles, les robots ont appris à associer des notions préjugés de beauté et de santé, ce qui conduit à coder des personnes plus foncées comme implicitement malsaines. Un projet de toute évidence inoffensif a créé des machines qui, si on leur en donnait la chance, pourraient fournir aux personnes à la peau plus foncée des analyses de santé inexactes entraînant des conséquences très réelles. En reconnaissant ces conséquences, c'est facile pour les développeurs de Youth Laboratories de réexaminer les données et d’écrire des algorithmes pour éviter cette conclusion, mais malheureusement cela ne

pourra pas suffire. La préférence implicite pour les blancs sera toujours reflétée si c'est un phénomène dans la société. Elle parle de cette situation délicate en l'appelant une époque de «New Jim Code» qui fait référence à l'époque de «Jim Crow» qui a rendu obligatoire la ségrégation raciale aux États-Unis. Le biais dans le code ne se présente pas seulement avec la défavorisation des noirs. En tant que société, on doit s'occuper de ce biais contre les groupes minorisés car les algorithmes le refléteront toujours. Ruha dit “Le danger de l'impartialité du New Jim Code est la négligence de l'iniquité persistante ... Dans ce contexte, les algorithmes ne peuvent pas être simplement un vernis qui couvre les lignes de faille historiques. Ils semblent également rationaliser la discrimination” (The danger of New Jim Code impartiality is the neglect of ongoing inequity... In this context, algorithms may not just be a veneer that covers historical fault lines. They also seem to be streamlining discrimination)[22]. Elle est consciente du fait que c'est impossible d'écraser les biais qui sont soutenus historiquement alors on doit prendre l'initiative de créer des audits où autres outils abolitionnistes[23] pour imposer une pratique de responsabilité alors que les technologies existantes et nouvelles satisfont le critère de l'équité dans le code.

De l'autre côté, en raison de la nature illégale de la collecte de statistiques raciales en France, trouver des études qui se concentrent sur les biais raciaux est presque impossible. Pour contourner cela et démontrer des biais raciaux dans l'usage de technologie, j'ai cherché des statistiques sur l'usage des appareils de reconnaissances faciales pour découvrir des motifs. Dès 2012, la France a institué le Traitement d'antécédents judiciaires (TAJ) créé par CGI une entreprise de conseil de l'informatique et en affaires. TAJ est un fichier utilisé par les policiers et

[22] Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, 2020.

[23] Le terme abolitionniste vient de la pratique de l'esclavage aux États-Unis. Les abolitionnistes étaient des gens qui voulaient mettre fin au système oppressif de l'esclavage. Dans ce contexte, les outils abolitionnistes sont des outils avec le seul but de détruire les manières dont la société promeut et soutient les systèmes oppressifs.

la gendarmerie nationale qui contient l'information personnelle et l'information biométrique de presque 19 millions de gens en France, à peu près une personne sur trois, qui ont déjà commis un crime[24]. Les policiers et la gendarmerie utilisent le TAJ pour les aider dans leurs enquêtes et trouver plus facilement les mises en cause. Un système comme le TAJ part du principe que les contrevenants précédents sont les plus probables de commettre de nouveaux crimes. Ce principe est problématique parce que cela implique déjà un biais contre les contrevenants précédents qui conduit les responsables à se focaliser sur eux en empêchant l'enquête. En plus de ce biais, le danger continue avec l'accès aux informations biométriques des citoyens. CGI se vante que son système améliore l'efficacité de la police en améliorant l'identification des suspects et les biens volés, fournissant une comparaison d'images pour identifier les personnes selon les physiques et des caractéristiques telles que les tatouages et des images faciales, ainsi que fournir des outils de surveillance économique pour les statistiques de la police[25]. Essentiellement, CGI permet aux policiers en France d'utiliser la reconnaissance faciale qui est une technologie développée par la vision numérique, un sous-champ de l'IA, pour identifier des suspects potentiels. Comme tout dans l'IA, la reconnaissance faciale est susceptible de biais et préjugés dans les algorithmes qui la soutiennent et les données qui lui sont fournies. Comme l'exemple des États-Unis, les systèmes d'IA peuvent introduire leurs subjectivités dans leurs travail qui peut blesser des gens, effectivement cette technologie dans les mains des policiers est dangereuse. Un documentaire américain qui s'appelle *Coded Bias*, traite des lacunes des systèmes de reconnaissance faciale actuels. Une chercheuse à MIT, Joy Buolamwini, a découvert que les systèmes de reconnaissance faciale largement disponibles ne fonctionnent pas sur les visages noirs. Ils ne sont

[24] CNIL. “TAJ : Traitement d'antécédents Judiciaires.” *CNIL*, 15 Nov. 2018, www.cnil.fr/fr/taj-traitement-dantecedents-judiciaires

[25] CGI. “Improving Performance with End-to-End TAJ Solution.” *CGI*, www.cgi.com/en/case-study/improving-performance-end-end-taj-solution

pas habitués à effectuer des analyses sur la peau plus foncée et en fait, ils ne réalisent pas la tâche de reconnaissance faciale sur les visages avec la peau plus foncée avec précision. Dans un environnement de recherche, des résultats imprécis ne sont pas graves mais aux mains de la police et la gendarmerie nationale, un résultat de reconnaissance faciale imprécis signifie que quelqu'un est accusé à tort d'un crime. Buolamwini argue que les algorithmes biaisées et imprécis blessent des gens ce qui est difficile à régler car on ne sait pas qui tenir responsable ou comment le faire. Dans ce cas en France, en plus de ne pas avoir de législation pour protéger les citoyens contre les décisions algorithmiques subjectives qui peuvent mettre la vie en danger, on ne peut pas défendre les gens avec la peau foncée contre les mêmes algorithmes prédateurs. Car le pays ne reconnaît pays la race comme un facteur distinctif et la race en France est étiquetée comme "données subjectives"[26], comment on gère la possibilité de ne pas identifier correctement des citoyens noirs comme des criminels? Aux États-Unis, après avoir découvert les défauts du système de reconnaissance faciale, Buolamwini a introduit de nouvelles données et algorithmes pour augmenter sa précision pour les gens avec la peau foncée. Si la France ne reconnaît pas la différence entre les gens avec la peau foncée et la peau claire, est-il possible d'avoir un système précis qui ne cible pas les groupes minoritaires?

Alors que la France n'est peut-être pas prête pour reconnaître la race comme un facteur distinctif, les autres pays francophones la reconnaissent et la discutent. En cherchant les attitudes de la France sur la présence du biais systémique dans les données qui sont reflétées dans les algorithmes de l 'IA, il se trouvait plusieurs études canadiennes sur ce biais racial et la façon de l'attaquer. Un rapport de l'Institut de recherche et d'informations socio-économiques (IRIS) à Québec traitait du racisme systémique dans la région.

---

[26] Norimitsu Onishi. "Les Idées américaines menacent-elles la cohésion française?" *New York Times*, 9 Feb. 2021

> En 2019, un rapport du Service de police de la ville de Montréal (SPVM) a montré que la probabilité moyenne de se faire interpeller par le SPVM variait en fonction de l'appartenance raciale. Les personnes noires et autochtones se font interpeller entre 4 et 5 fois plus que les personnes blanches. Les femmes autochtones ont quant à elles 11 fois plus de chances de se faire interpeller que les femmes blanches. Le constat saisissant de ce rapport est ainsi que les personnes noires et les femmes autochtones se font interpeller de manière disproportionnée par rapport à la taille de leur population[27].

Le rapport reconnaît l'aspect racial de ces interpellations et confirme leur caractère injuste. En reconnaissant cela, le gouvernement et les entreprises à Québec continuent à développer les manières dont ils peuvent agir pour tenter de résoudre ce problème et s'assurer que les personnes minorisées sont protégées. Pour conclure, le rapport dit "en clair, ce n'est qu'en reconnaissant l'existence du racisme systémique que le gouvernement pourra lutter efficacement contre ses diverses manifestations"[28]. L'acte de reconnaissance est le premier pas pour combattre ces biais si prévalents et c'est un acte devant lequel la France recule. Comme le Canada avait des racines francophones, c'est intéressant d'examiner l'attitude sur le biais systémique d'un pays entre le monde francophone et américain. Pendant que le Canada a sa propre histoire méprisable contre les groupes minorisés, on voit que les canadiens collectent les statistiques raciales et, conséquemment peuvent reconnaître des motifs contre les gens qui sont ciblés systématiquement. Cette volonté d'admettre qu'il y a un problème qui cible des groupes minorisés et de reconnaître qu'il faut le régler pour les protéger distingue le Canada de la France. Marine le Pen et d'autres en France ont un dédain pour cette culture canadienne en disant qu'elle a "baigné dix ans dans la culture américaine"[29]. D'après eux, il apparaît que les idées américaines ont déjà menacé le Canada. On voit un rapprochement du Canada aux idéologies américaines malgré son histoire francophone. Peut-être est-ce à cause de la reconnaissance des conséquences de l'histoire

---

[27] Wissam Mansour, et al. *Qu'est-Ce Que Le Racisme Systémique?* 4 June 2020, https://iris-recherche.qc.ca/blogue/qu-est-ce-que-le-racisme-systemique

[28] IBID

[29] Norimitsu Onishi. "Les Idées américaines menacent-elles la cohésion française?" *New York Times*, 9 Feb. 2021

commune des États-Unis et du Canada du génocide des peuples autochtones; on ne sait pas. Il y a de nombreux événements dans l'histoire qui se cumulent pour créer les cultures et les croyances des sociétés actuelles et les subjectivités des principes fondateurs de ces sociétés seront toujours reflétés dans les aspects de celle-ci aujourd'hui. À mon avis, la France stagne dans son approche d'ignorer des statistiques raciales importantes. Je comprends la raison pour laquelle cette pratique est significative mais pour que la société avance au-delà d'un héritage dépassé et pour attaquer les vrais menaces sociales, il faut mettre à jour les pratiques standard. Peut-être les idées américaines ne menacent pas la cohésion française, peut-être qu'elles contestent les normes françaises et exposent les défauts sociétaux. Malgré leurs approches évidemment différentes dans la construction de la société, les principes fondateurs et les attitudes actuelles à l'égard de la race, la France et les États-Unis ont atteint le même objectif: la discrimination contre les groupes minorisés reflétée dans la technologie.

Conclusion

L'IA est si commune que le citoyen moyen ne sait pas qu'il interagit avec les systèmes de l'IA même si c'est une activité quotidienne. Le champ de l'IA est en train d'être industrialisé et en 2018, "le marché de l'intelligence artificielle ne représente ... que 4 milliards de dollars, son expansion est telle qu'il devrait avoisiner les 60 milliards d'ici 2025 selon les spécialistes"[30]. On ne peut pas échapper à la prévalence de l'IA dans nos sociétés, alors on doit construire une société où le pouvoir de l'IA est contrôlé. On a vu plusieurs exemples de la puissance de la technologie et aussi de ses déficiences. Même si cette technologie est controversée, elle devient de plus en plus répandue dans les sciences, les grandes entreprises et même le gouvernement.

[30] "IA : Les Chiffres à Connaître En France Et Dans Le Monde: Comarketing-News." *Comarketing-News*, 12 July 2018, https://comarketing-news.fr/ia-les-chiffres-a-connaitre-en-france-et-dans-le-monde/

Selon la façon dont la technologie est utilisée, l'IA pourrait changer le monde pour le meilleur ou pour le pire comme on l'a examiné. C'est à nous de décider. En gardant en tête les enjeux éthiques de lieu d'autorité et les dilemmes moraux, on peut assurer que l'IA est utilisée de la manière la plus responsable possible. On doit avoir un corps consultatif pour réduire des conflits d'intérêts des développeurs et des chercheurs. Le corps consultatif aussi doit s'assurer que tous les projets et toutes les initiatives sont pour le bien public et ont considéré les implications éthiques. On doit aussi composer des équipes de recherches diverses qui reflètent des perspectives différentes pour gérer les biais implicites et explicites. La diversité et la collecte des expériences et points de vue de différents membres de l'équipe limitent les types de biais qui pourraient se produire. Même après avoir suivi tous les conseils éthiques, c'est presque impossible d'éliminer les biais contre les groupes minorisés à cause du biais systémique.

Pour réduire les conséquences de ce biais, les ingénieurs et développeurs doivent réexaminer les données et les algorithmes qu'ils utilisent pour signaler tout ce qui aurait la possibilité d'être nuisible. Aussi, il faut que les autres personnes hors de l'équipe de recherches puissent évaluer les données et algorithmes d'une perspective plus objective. C'est possible de trouver les biais dans les données et de changer les algorithmes pour contrer activement les biais, mais la meilleure façon de réduire les biais dans les données est de produire des données moins biaisées. On ne peut que produire des données moins biaisées en progressant en tant que société et combattre contre les principes fondateurs qui ont soutenu ces préjugés. S'il n'y a pas de progrès sociétal, l'IA va toujours refléter les biais qu'il trouvera. En attendant, on ne peut qu'être conscient de la puissance nocive que l'IA donne à la société et développer les méthodes pour assurer que nos programmes sont audités et vérifiés avec la plus grande prudence possible. Une façon de faire cela est de changer la manière dont l'IA est commercialisée. Maintenant, la plupart

des commercialisations de l'IA promeut un système omniscient et objectif qui ne reflète pas de subjectivités. On a vu que ce n'est pas le cas. En fait c'est l'opposé. En France et aux États-Unis, les systèmes de l'IA reflètent leurs biais sociétaux respectivement et les amplifient. Alors qu'un changement systémique profond n'est pas encore observé dans l'une ou l'autre de ces sociétés, on peut commencer à gérer les enjeux éthiques de l'IA avec des manières que j'ai mises en avant. Ce ne sont pas des solutions permanentes, mais un bouche-trou pendant qu'un travail important est effectué pour reconnaître proprement les conséquences du biais systémique et développer adéquatement la structure afin de soutenir les groupes minorisés affectés par ce biais.

Bibliographie

Ayoh, Christine. “L'Automatisation d'un robot social .” *Institute for Field Education*, 2019.

Bender, Emily M., Gebru, Timnit, McMillan-Major, Angelina, and Mitchell, Margaret. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3442188.3445922

Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, 2020.

Bradley-Popovich, Greg E. “An Exercise in Ethics: Case Analysis with Implications for the Exercise Physiologist .” *Professionalization of Exercise Physiology*, vol. 3, no. 6, 2000, https://www.asep.org/asep/asep/ExerciseInEthics.html.

CGI. “Improving Performance with End-to-End TAJ Solution.” *CGI*, www.cgi.com/en/case-study/improving-performance-end-end-taj-solution

Coded Bias. Directed by Shalini Kantayya, appearances by Joy Buolamwini, Meredith Broussard, and Cathy O'Neil, 7th Empire Media, Chicken And Egg Pictures, Ford Foundation - Just Films, ITVS, Women Make Movies, 2020.

CNIL. “TAJ : Traitement d'antécédents Judiciaires.” *CNIL*, 15 Nov. 2018, www.cnil.fr/fr/taj-traitement-dantecedents-judiciaires

COMEST. (2019). Étude préliminaire sur l'éthique de l'intelligence artificielle. In *World Commission on the Ethics of Scientific Knowledge and Technology*. Paris, Île-de-France.

Comité d'éthique du CNRS. *Pratiquer une recherche intègre et responsable,* Comité d'Éthique du CNRS, 2017. https://comite-ethique.cnrs.fr/wp-content/uploads/2019/10/GUIDE-2017-FR.pdf

“The Declaration of Independence.” *National Archives and Records Administration*, National Archives and Records Administration, www.archives.gov/founding-docs/declaration

Demiaux, Victor. “Comment permettre à l'homme de garder la main ?” CNIL, 2017, www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf

Devillers, Laurence (2019). “Dimensions affectives et sociales des interactions parlées avec des (ro)bots et enjeux éthiques”, Limsi.fr, https://www.limsi.fr/fr/recherche/tlp/themes/dimensions-affectives-et-sociales-des-interactions-parlees , 17 novembre 2019.

Gordon, John-Stewart, and Sven, Nyholm. “Ethics of Artificial Intelligence.” *Internet Encyclopedia of Philosophy*, ISSN 2161-0002, https://iep.utm.edu/ethic-ai/ .

Hao, Karen. “We Read the Paper That Forced Timnit Gebru out of Google. Here's What It Says.” *MIT Technology Review*, MIT Technology Review, 7 Dec. 2020, www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

History.com Editors. *The Holocaust*. 14 Oct. 2009,

www.history.com/topics/world-war-ii/the-holocaust

"IA : Les Chiffres à connaître en France et dans le monde: Comarketing-News." *Comarketing-News*, 12 July 2018, https://comarketing-news.fr/ia-les-chiffres-a-connaitre-en-france-et-dans-le-monde/

"Jean-Gabriel Ganascia." *CNRS*, www.cnrs.fr/en/person/jean-gabriel-ganascia.

"Life, Liberty and the Pursuit of Happiness." *Wikipedia*, Wikimedia Foundation, 19 Apr. 2021, https://en.wikipedia.org/wiki/Life,_Liberty_and_the_pursuit_of_Happiness

Mansour, Wissam, Forcier, Matthieu, Posca, Julia, Couturier, Eve-Lyne, Langevin, Raphaël, Hurteau, Philippe & Hébert, Guillaume. *Qu'est-Ce Que Le Racisme Systémique?* 4 June 2020, https://iris-recherche.qc.ca/blogue/qu-est-ce-que-le-racisme-systemique

Kendi, Ibram X. *Stamped from the Beginning: the Definitive History of Racist Ideas in America*, Bold Type Books, 2017, pp. 1–11.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2018.

Onishi, Norimitsu. "Les Idées américaines menacent-elles la cohésion française?" *New York Times*, 9 Feb. 2021, www.nytimes.com/fr/2021/02/09/world/europe/France-universites-decolonialisme-woke.html

Seibt, Sébastian. "Timnit Gebru, licenciée par Google pour sa vision 'éthique' de l'intelligence artificielle?" *France 24*, France 24, 8 Dec. 2020, www.france24.com/fr/%C3%A9co-tech/20201208-intelligence-artificielle-la-question-%C3%A9thique-qui-a-fait-tiquer-google

Tyrrell, Ian. "What, Exactly, Is 'American Exceptionalism'?" *The Week - All You Need to Know about Everything That Matters*, The Week, 21 Oct. 2016, https://theweek.com/articles/654508/what-exactly-american-exceptionalism

Villani, Cédric. *Donner un sens à l'intelligence artificielle: pour une stratégie Nationale européenne: Mission Parlementaire du 8 septembre 2017 au 8 mars 2018*, 2018. https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf