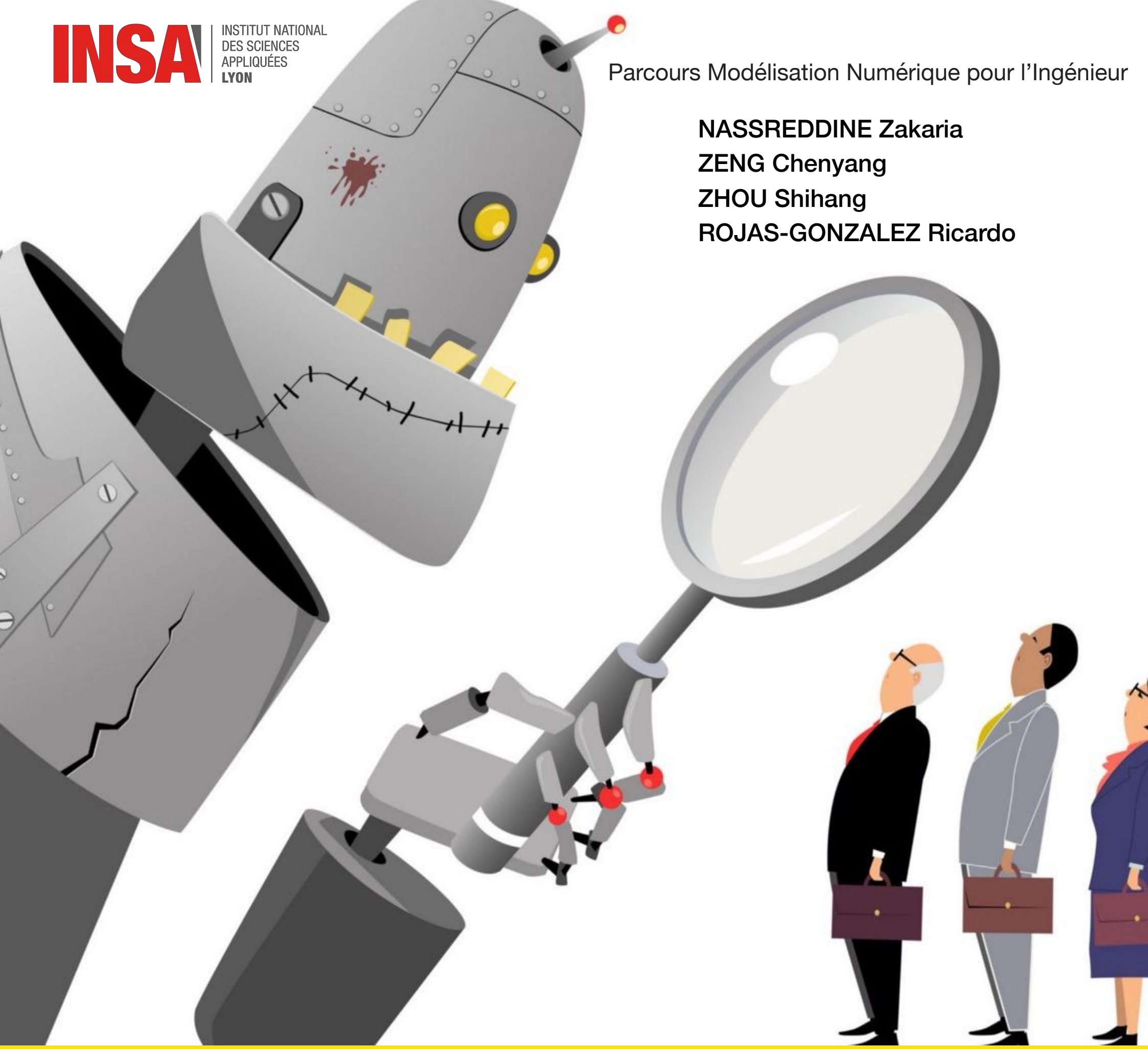
INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

Parcours Modélisation Numérique pour l'Ingénieur

NASSREDDINE Zakaria
ZENG Chenyang
ZHOU Shihang
ROJAS-GONZALEZ Ricardo



La discrimination à l'ère des Big Data : Intelligence artificielle et amplification de l'oppression systémique

L'intelligence artificielle gagne en omniprésence à l'ère où la taille de nos données devient astronomique, elle est là pour appuyer nos décisions, trier nos populations et évaluer nos performances. Considérés pendant longtemps comme étant le paroxysme de l'objectivité, ces modèles informatiques sont de plus en plus critiqués et accusés de 'menacer la démocratie et d'accroître les inégalités'[(1)].

Tout comme au niveau de la société humaine, la partialité a envahi l'intelligence artificielle.

Ceci peut être expliqué, dans un premier temps, par la présence de biais intériorisés chez la personne ayant conçu et entraîné le modèle. Le problème surgit au moment où ces derniers produisent des prédictions et appuient des décisions motivées, entre autres, par des attributs controversés, à savoir le genre, l'ethnie, la classe sociale, la religion…etc.

Sans avoir à être explicitement dépendants de tels facteurs, plusieurs modèles ont présenté différents taux de favoritisme. C'est avec des années de recul et de longs périples de désillusion que des chercheurs, autrefois conditionnés à ne jurer que par ces algorithmes rigoureusement équitables et dépourvus du moindre motif pour discriminer, se rendent compte que leurs inventions, a priori des concentrés de pure logique et de raisonnement rationnel ont appris à aggraver les mêmes injustices qu'ils ont été mis en place dans le but de neutraliser.

Que cela en soit l'intention ou non, les modèles basés sur les Big Data s'avèrent capables de servir d'outils d'oppression. On s'intéressera surtout aux cas de figures où c'est la modélisation qui fait défaut, mais il demeure important de noter que dans d'autres situations, il s'agit bien d'une volonté de se servir d'outils numériques pour maintenir un status quo privilégiant certains groupes de façon délibérée.

À LA MERCI DES MATHS

De nos jours aux États-Unis, quelques 72% de CVs n'ont jamais la chance d'être évalués par un être humain. L'utilisation abondante de ces systèmes de tri basés sur des formules mathématiques compliquées, accompagnée d'un manque alarmant de transparence réduit incroyablement le nombre de candidats au courant des raisons pour lesquelles ils n'ont pas réussi à passer le cap. Dans le rare cas où on l'apprend, très peu sont prêts à se retourner vers un consultant juridique. Ces algorithmes tournent en permanence, en fouillant des pétaoctets d'informations à la recherche de modèles et de corrélations s'intéressant de moins en moins à des marchés financiers et de plus en plus à des êtres humains. Nous en sommes les nouveaux sujets. Mathématiciens et statisticiens analysent nos actions, nos désirs, à la transaction près, pour pouvoir prédire notre potentiel en tant qu'employés, étudiants, délinquants ou amants, en baignant dans l'éloge du grand public bien rassuré par l'objectivité apparente de ces outils révolutionnaires. Entre des juges prononçant des peines plus sévères à l'heure du déjeuner sous l'effet de la faim et des locaux quasi-exclusivement remplis des neveux du manager, le recours à des méthodes éliminant ces subjectivités a été amplement légitime, d'autant plus lorsqu'ils ont le mérite d'être rigoureusement scientifiques. C'est pourtant là où réside le problème. En effet, étant conçus par des humains faillibles de nature, ces algorithmes ne sont pas à l'abri de reproduire des préjugés et de chiffrer des idées reçues dans des modèles opaques. Tels des dieux, leurs modes de fonctionnement sont invisibles au commun de la population, à l'exception des grands prêtres de leur domaines: informaticiens et mathématiciens. Un fois ils arrivent à un verdict, aussi erroné et nuisible qu'il puisse être, il n'y a plus moyen qu'il soit remis en question, et ces décisions ont tendance à punir les pauvres et les tranches les plus marginalisées de la société, tout en assurant la prospérité de la poignée en haut de la pyramide. À l'heure où des algorithmes similaires mis en place par de grandes franchises de football pour évaluer des joueurs valant des millions de dollars sont constamment revus et ajustés, le tri de la populace met un jeu un risque moins important; on s'en moque du devenir du troupeau éliminé.

FOCUS: COMPAS OU LA POLICE PRÉDICTIVE AFROPHOBE

Dans une tentative de réduire les taux d'incarcération à travers les États-Unis et d'alléger la charge des institutions d'accueil du système juridique pénal, plusieurs états ont mis en place un système d'évaluation des risques qui fournit un score indiquant si un individu arrêté par les forces de l'ordre présenterait ou non une forte tendance à la récidive criminelle. Ce score est destiné à informer le juge sur le danger potentiel que représente cet individu pour sa communauté, avant de pouvoir se prononcer sur la possibilité de libération conditionnelle ou, le cas échéant, la durée d'emprisonnement. Parmi les données collectées sur les défendants, extraites de leurs casiers judiciaires ou en demandant directement aux accusés de répondre à des questions, aucune ne rendait compte de leurs groupes ethniques, ou du moins pas de manière explicite. Une enquête menée par ProPublica montre que les noirs étaient deux fois plus susceptibles d'être, à tort, déclarés à risque élevé et ne plus se faire arrêter de nouveau, alors que les blancs avaient deux fois plus de chance d'être identifiés comme présentant un faible risque pour ensuite récidiver en sortant[(2)]. Face à ces accusation, Northpointe, entreprise ayant développé cet outil dont le siège se trouve au Michigan, refuse de rendre public son algorithme pour des raisons de propriété économique et intellectuelle. Les accusations persistent, même si le modèle ne prenait pas en variable d'entrée l'ethnie, cela ne lui empêchait pas de retenir d'autres données qui y sont fortement corrélées : en enregistrant le statut marital des parents par exemple, dans un pays où les familles afro-américaines sont disproportionnellement plus concernées par le divorce, il se retrouvait en train d'encoder involontairement cet attribut racial. Plein d'autres paramètres permettent une codification implicite de l'ethnie dans un contexte où l'activité policière est déjà concentrée au niveau des communautés hébergeant plus de minorités. Le remède semble difficilement atteignable si l'on ne veut renoncer à toutes ces données, quitte à gagner en équité au détriment de la précision.

Bernard PARKER. Score : 10 (risque élevé). Infractions antérieures : résistance non violente à une arrestation. Infractions ultérieures : aucune.

Dylan FUGETT. Score : 3 (faible risque). Infractions antérieures : tentative de cambriolage. Infractions ultérieures : 3 possessions de drogue.

DE LA FUITE À L'APPRENTISSAGE DES PRÉJUGÉS

Dans un premier temps, on pourrait bien établir le résultat suivant : une des sources d'implémentation de biais dans un algorithme est la façon doit on le conçoit qui découle de la manière dont on pense la société. À un moment donné, il faut bien partir de sa propre vision de ce que serait un bon candidat, un client de confiance, un délinquant méritant une seconde chance. Mais on reste bien loin du diagnostic exhaustif.

'En améliorant les échantillons, on peut éviter les erreurs'

'Comparaison des performances et individus quantifiables, quelque part, il y'a quelque chose d'aberrant, de complètement déshumanisant' nous disait MIHARA-TEYSSIER Norio, derrière son bureau d'enseignant-chercheur au département des humanités de l'INSA de Lyon, abordant d'un regard sceptique cette numérisation ubiquiste de la société. Au-delà de la philosophie même d'un modèle, il souligne l'urgence de la dimension statistique vitale à son fonctionnement. 'Les échantillons ne sont pas bien équilibrés, en portant une attention particulière à la qualité des données fournies au modèle, on peut s'affranchir de ces erreurs.' Cette question de pertinence des données et de leur capacité à représenter de manière fidèle la société s'avère absolument cruciale. Même en en arrivant là, on serait bien naïf de croire que la solution serait aussi élémentaire que de se limiter à surveiller la fidélité des data. Le problème, perçu dans un autre contexte comme atout majeur, de l'apprentissage machine est la compensation des données. Lorsque des attributs non pertinents (comprendre: discriminatoires) sont vivement corrélés à des attributs clés, ils seront fortement pondérés par le modèle qui apprendra à amplifier le rôle qu'ils jouent dans le tri. En omettant l'intégralité des paramètres condamnables, on n'élimine que les biais 'directs', et selon la manière dont c'est fait, on risque de réintroduire et de renforcer les biais implicites.[3] La réalité étant elle même injuste, des modèles aveugles à ce qui motive ces injustices ne se peuvent d'être équitables. Dans une société conditionnée par des siècles d'antécédents de privilège et régie par une ensemble de schémas récurrents de pouvoir, un modèle insensible à ce contexte est condamné à non seulement maintenir, mais amplifier ces iniquités.

EST-CE PEINE PERDUE ?

Le processus de la mise en place d'une IA dans un contexte social est plus analogue à l'éducation d'un enfant qu'à l'implémentation d'une simple application. Il faut garder en tête que c'est destiné à interagir avec des humains et avoir un impact direct, éventuellement irréversible, sur leurs vies. Darin Stewart, vice-président de recherche chez Gartner Analyst, tient à mettre l'accent sur l'interprétabilité des modèles, en leur empêchant d'être libres de détecter toute corrélation statistiquement pertinente et incite les spécialistes à entraîner leurs algorithmes à la maîtrise des subtilités des contextes sociaux dans lesquels ils évoluent.[4] Il s'agit d'un nouveau besoin en profils plus transversaux et mieux sensibilisés à ces enjeux.

'Bien développer une intelligence artificielle, c'est élever un bon citoyen'

3 QUESTIONS À

Docteure FAVRE Cécile
Maîtresse de conférences en informatique et chercheuse associée du laboratoire de sociologie CMW.

Quels enjeux éthiques se rapportent au choix de déléguer à une intelligence artificielle le pouvoir d'évaluer et de juger des populations ?

Il est important de bien cerner l'état d'esprit dans lequel on fait ces algorithmes et d'avoir conscience que cela se passe dans un environnement social qui fait que nos choix de sujets ne sont pas anodins. Ce n'est pas pour autant parce qu'il y'a des gens qui travaillent ces questions de l'éthique à temps plein que dans le domaine de l'informatique on ne doit pas se remettre en question. Soit en collaborant et en discutant mais aussi en nourrissant nos propres réflexions. L'éthique doit être pensée d'un point de vue de la composition de la société. Le grand enjeu pour moi, c'est de vraiment se poser la question et de dépasser la croyance que parce que c'est des machines, des algorithmes et que c'est artificiel, que c'est neutre.

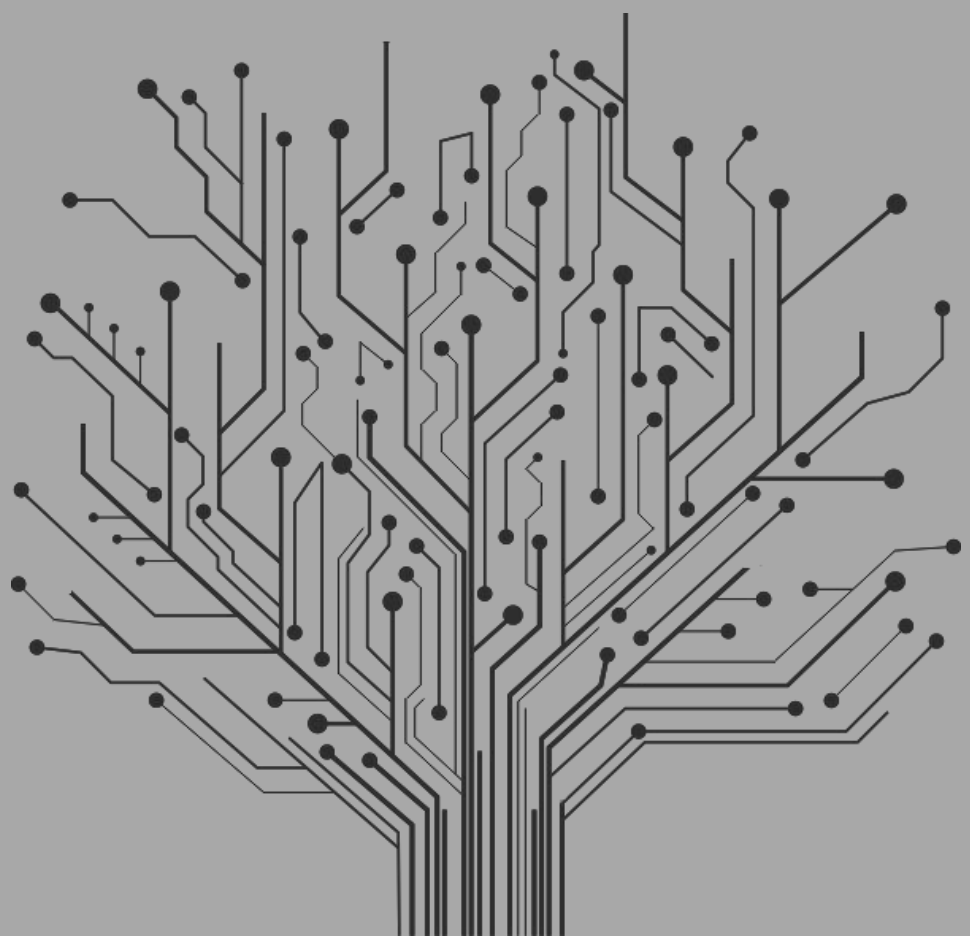
'Il faut dépasser la croyance que c'est parce que c'est artificiel, c'est systématiquement neutre [...] et réfléchir plus profondément à ce qu'on veut comme intelligence, comme société.'

Quelle place donnez-vous aux sciences humaines et sociales dans l'accompagnement de la construction de ces modèles ?

Accompagner l'IA dans une bonne direction sociale, c'est surtout penser les programmes (en sciences et techniques) de demain et déconstruire les modèles actuels selon lesquels on fait de la recherche. Il ne s'agit pas forcément de former exclusivement des profils à double casquette, mais d'encourager l'acquisition de compétences transversales, et d'assurer l'ouverture de ces formations, pour la plupart cloisonnées, aux enjeux sociaux. Les sciences humaines et sociales ont beaucoup à apporter de ce point de vue, mais se retrouvent malheureusement délaissées, principalement en terme de financement, vu la hiérarchie qui s'est construite dans le paysage des sciences et de la recherche. Pour arriver à se faire comprendre entre deux domaines où on ne parle le même langage, il faut une réelle collaboration qui doit commencer par une vraie valorisation des SHS, vitales à ces enjeux.

Quelle est votre vision d'un modèle juste dans une optique de prise de décision à fort impact humain ?

Il faut voir cela sur deux volets indissociables et qui travaillent ensemble : la partie programme ou algorithme et la partie données. Même quand on a un programme qui se veut aveugle à la question du genre par exemple, mais que dans les données la trace de cette variable est implicitement présente, elle peut potentiellement être utilisée. Cette vision de neutralité ou de justice repose sur ce qu'on veut collectivement comme société. En utilisant les données actuelles, la société n'étant pas égalitaire, on risque de reproduire et d'accentuer encore ces disparités. C'est à ce moment là où on pourrait s'intéresser à des contraintes d'apprentissage, on parle souvent d'action positive, pour rattraper ces injustices et rééquilibrer la réalité. On se retrouve donc face à des enjeux politiques conséquents, dans une société où même quand on n'est pas contre plus d'égalité, on demeure réticent à perdre des privilèges.



RESSOURCES

Interviews :

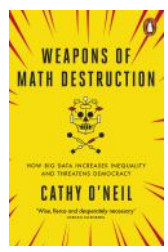
Cécile FAVRE

Maîtresse de conférences en informatique, Université Lumière Lyon 2
Membre du laboratoire d'informatique ERIC
Chercheuse associée du laboratoire de sociologie CMW
Rattachée à l'UFR Anthropologie, Sociologie, Science Politique
Co-responsable de la Mention de Master Etudes sur le Genre de Lyon

Norio MIHARA-TEYSSIER

Professeur de japonais et de sciences humaines à l' INSA Lyon.

Bibliographie :



[1] Cathy O'Neil. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Édition Crown (6/09/2016), 272 pages, ISBN-13: 978-0553418811.

[2] ProPublica. Machine Bias: Risk Assessment in Criminal Sentencing. Disponible sur : https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (Consulté le 22/04/2019 de Villeurbanne)

[3] TowardsDataScience. Preventing Machine Learning Bias. Disponible sur https://towardsdatascience.com/preventing-machine-learning-bias-d01adfe9f1fa (Consulté le 01/05/2019 de Villeurbanne)

[4] TechTarget. How bias in AI happens -- and what IT pros can do about it. Disponible sur : https://searchcio.techtarget.com/news/252447877/How-bias-in-AI-happens-and-what-IT-pros-can-do-about-it (Consulté le 26/04/2019 de Villeurbanne)

FranceCulture - Les chemins de la philosophie L'intelligence artificielle a-t-elle du coeur ? Emission disponible sur : https://podcasts.apple.com/fr/podcast/les-chemins-de-la-philosophie/id390165399?ign-mpt=uo%3D4 (Écouté le 22/05/2019 à Villeurbanne).

Images :

Photo de couverture (page 1) URL : https://medium.com/@turalt/ai-isnt-biased-we-are-b74ec94d1698

Image (page 3) URL : https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Photo de madame FAVRE (page 4) fournie par madame Cécile Favre elle-même.