



Le 14 décembre 2020

LE POUVOIR DES ALGORITHMES QUI NOUS GOUVERNENT

Nous remercions chaleureusement Ahmed Mehaoua pour son accueil à Telecom Paris, Astrid Bertrand, David Bounie et Winston Maxwell, chercheurs au sein du Laboratoire de Telecom Paris qui ont présenté leurs travaux sur la gouvernance des algorithmes, dans le cadre des « Lundi de la Cybersécurité » du 7 décembre 2020. Comment la gouvernance est-elle exercée et par qui ?

Deux catégories d'algorithmes sont présentées, après avoir rappelé qu'un algorithme est une suite d'opérations ou d'instructions permettant d'obtenir un résultat.

Certains algorithmes participent à l'aide à la décision, d'autres sont totalement autonomes. Parmi ceux qui apportent une aide dans la prise de décision, nous allons trouver par exemple les algorithmes qui prédisent des délits en fonction des quartiers, des vols... Ces algorithmes sont dédiés à la justice, la police.

D'autres algorithmes sont autonomes et permettent de fournir des services automatiques de vidéo surveillance ou des prestations dans l'industrie.

Nos chercheurs se penchent sur les algorithmes dits « non autonomes ».

Leurs travaux s'appuient sur cinq axes de recherche : l'intérêt général, les biais et l'équité, l'explicabilité, la gouvernance et la régulation, la responsabilité.

C'est en tenant compte du droit, de l'utilisateur, des finalités de l'algorithme et des tendances jurisprudentielles que la recherche tente de comprendre comment l'intelligence artificielle peut et doit se développer, afin de faciliter l'acceptation et l'utilisation d'algorithmes par la population.

1 - De l'algorithme opaque à l'algorithme transparent.

L'algorithme doit pouvoir apporter le meilleur résultat, la meilleure décision ce qui amène nos chercheurs à détecter les limites de l'outil, les faux positifs tout comme les biais d'équité et de discrimination.

Il est impossible pour un algorithme de répondre à toutes les situations

La question posée est de savoir si une règle mathématique pourrait répondre à toutes les situations et garantir l'équité. L'équité est un concept éthique qui se traduit en mathématiques par une recherche d'équité algorithmique.

Pour chacun des domaines d'application des algorithmes, (lutte contre le blanchiment), Nos chercheurs s'intéressent aux ratios, aux différences (par exemple, Hommes – Femmes).

Beaucoup trop d'hypothèses sont à prendre en compte pour traduire et mesurer l'équité en termes statistiques pour refléter des équités, sans tomber dans un biais d'équité (genre, origine ethnique).

Les différences entre les groupes doivent être les plus faibles possibles. Entre les données d'entrée et le résultat, une série de biais sont analysés parmi lesquels nous trouvons les biais d'équité.

Les biais d'équité doivent être pris en compte, afin d'éliminer toute pratique discriminatoire. Ces biais ne sont pas identiques s'il s'agit de personnes ou de groupes et entre les groupes, les résultats sont différents, certains groupes ou certains individus étant eux-mêmes protégés par une loi.

Un poids est attribué à certains paramètres pour éviter des discriminations ou pour intégrer une inégalité (ex des impôts plus forts pour les foyers les plus riches).

Il n'est pas possible dans les calculs mathématiques de prendre en compte à la fois l'individu et le groupe dans les analyses menées.

Nos chercheurs nous font observer que l'utilisation des algorithmes dans certains domaines tels que la justice prédictive permettrait aujourd'hui d'observer qu'un algorithme prédit correctement 61% des cas de récidive parmi 10000 condamnés aux USA, tout en revisitant les hypothèses initiales.

Entre les récidives parmi les populations blanches et les récidives parmi les populations noires, les récidives parmi les populations noires ayant été surestimées, alors que les récidives parmi les blancs auraient été sous estimées.

La question fait également débat tant il paraît compliqué de traduire sous une forme mathématique des concepts juridiques qui protègent les droits fondamentaux, d'une part et alors que le droit (secret des affaires, par exemple) pourrait également être invoqué pour protéger la conception de l'algorithme.

Les contraintes légales font obstacle au pouvoir absolu de l'algorithme

20 critères /attributs sont protégés par la loi ce qui complexifie la traduction de l'équité dans un algorithme sans aboutir à des discriminations.

L'équité algorithmique dépend également du contexte économique et politique. La traduction de l'équité en équité algorithmique n'est pas simple, l'un étant un concept éthique, alors que l'autre est un concept mathématique. La mesure mathématique de l'équité nécessite de faire des choix qui relèvent de la sphère politique.

Dès la conception de l'algorithme, il faudrait orienter sa conception en construisant nativement un outil doté d'une éthique. L'explicabilité de l'algorithme, sa transparence seront analysées par les régulateurs et les tribunaux. Encore faut-il pouvoir construire et offrir une explicabilité satisfaisante de l'algorithme.

2- L'explicabilité des algorithmes

Comment aboutir à une explicabilité satisfaisante des algorithmes auprès des utilisateurs ?

L'explicabilité de l'algorithme, est une information donnée à l'Utilisateur sur le fonctionnement de l'algorithme dans son ensemble (explication globale) et une information sur une décision en particulier (explication locale).

Cette explicabilité permettrait d'augmenter la confiance des utilisateurs tout en protégeant leurs droits. Cette explicabilité doit servir l'exercice de test et de contrôle post réalisation de l'algorithme.

Aboutir à une explicabilité satisfaisante

Cette explicabilité doit permettre de répondre à un questionnement de l'Utilisateur, rester objective, compréhensible. La fiabilité de l'explication dépendra du receveur, des contraintes de temps, du contexte réglementaire, de la gravité de la décision.

L'Utilisateur et sa capacité d'interprétation

Du point de vue de l'utilisateur, comment donner du crédit aux résultats d'un algorithme, si le fournisseur de l'algorithme ne sait pas expliquer les résultats ? ou comment ne pas lui imputer une responsabilité algorithmique en cas d'erreur ?

Le sentiment d'opacité de l'algorithme n'est pas seulement le fait de l'algorithme. Nos chercheurs puisent leurs sources dans les sciences sociales. Les sciences sociales pointent les biais cognitifs qui viennent influencer la prise de décision et son interprétabilité.

L'Utilisateur, qu'il soit un individu ou un groupe a sa propre appréciation de l'acceptabilité, ses propres freins et malgré tout, le souhait de faire confiance à un algorithme dont la finalité est d'apporter un résultat, une aide dans une prise de décision.

Il y a autant de causes de X que d'explications de X.

Nos chercheurs nous rappellent les quatre causes d'Aristote. Un modèle d'IA génère une explication qui fait référence aux causes et aux chaînes de causes d'un événement qu'il s'agira d'isoler. On cherche à formuler des hypothèses. L'utilisateur pourra les interpréter.

Nos chercheurs vont s'évertuer à comprendre comment réduire l'écart entre « l'outil », l'Utilisateur, et le droit.

Après avoir rappelé qu'il n'y a pas un modèle d'algorithme mais des algorithmes pour chacune des finalités qu'il sert et que l'algorithme doit respecter l'équité, l'approche décrite par nos chercheurs consiste à utiliser l'algorithme également comme un moyen d'interaction au service de l'Utilisateur pour éclairer l'Utilisateur, lui donner confiance dans l'outil et lui permettre d'intégrer des informations qui proviennent de conseils algorithmiques.

« C'est l'autonomie de l'Utilisateur qui est recherchée. »

Ils explorent les pistes pour mesurer l'autonomie de l'Utilisateur, à travers :

- des explications pouvant permettre à l'Utilisateur de s'appropriier les explications;
- des recommandations quand l'Utilisateur a déjà sa propre réponse;
- les messages d'alerte ;

L'Utilisateur a lui-même une participation active dans l'appréciation du résultat.

3 - Mesurer et contrôler la transparence de l'algorithme

Alors comment respecter dans le même temps le socle de lois, de règlements, les Directives, la Charte Européenne des Droits de l'Homme qui encadrent nos droits fondamentaux et toutes les hypothèses de travail décrites ?

Il faudrait, en toute rigueur, éliminer les faux positifs, ce qui se révélerait être finalement une erreur, sans contrôle préalable du résultat.

Un contrôle externe par un tiers devra donc quant à lui être mené afin d'analyser le seuil de faux positifs qui lui-même sera apprécié différemment en fonction de la finalité attendue.

L'exemple est donné d'un algorithme qui est déployé pour analyser des données de trafic aérien et de localisation, pour surveiller les comportements suspects, afin de lutter contre le terrorisme.

Les algorithmes qui traitent des données de vols présenteraient 90% de faux positifs comme dans les activités de lutte contre la fraude menées par les banques. Et pourtant l'algorithme remplit parfaitement sa finalité, à savoir viser l'intérêt général en évitant un risque d'attentat.

Le contrôle humain prendra le relais (post conception), afin de procéder à une analyse critique du résultat.

Transparence et intérêt général

La question reste de savoir si l'intérêt général peut faire obstacle à toute obligation de transparence. La réponse est négative. En revanche, le traitement systématique des données est forcément contraire à la Charte Européenne des Droits de l'Homme. La conservation des données de trafic poserait également problème.

C'est sur la base d'un ensemble de textes (Loi Lemaire, Loi Informatique et Libertés, RGPD,...) qu'il faut apprécier les effets des règles qui encadrent les algorithmes.

Le Code des relations entre le public et l'administration établit également des obligations de transparence des algorithmes, en relation avec les administrés qui sont utilisateurs des services publics.

4 - Le Contrôle de la qualité de la transparence vu par la CJUE

La Cour de Justice de l'Union Européenne a eu l'occasion de se prononcer sur les conditions nécessaires pour assurer la transparence des algorithmes.

Il faut assurer :

- une adéquation entre le niveau d'ingérence causé par les algorithmes et l'objectif d'intérêt général. C'est l'équilibre, la proportionnalité entre l'Intérêt général, la sécurité nationale versus les moyens intrusifs qui sont appréciés. Les moyens intrusifs doivent s'exercer sur une période limitée. Le poids des justifications de l'intérêt général doit être documenté.

- un contrôle institutionnel ;

- une transparence suffisante pour permettre un contrôle effectif ;

Cela suppose que le modèle d'algorithme ait été préétabli dans cet objectif et soit spécifique à la finalité visée pour pouvoir effectuer le contrôle de transparence à posteriori.

- une fiabilité suffisante, (absence de discrimination et une intervention humaine avant la prise de décision).

Cela suppose que la démonstration des tests effectués doit avoir été anticipé, tracé, documenté.

Le modèle ne doit pas être discriminatoire mais explicable. C'est l'analyse des textes légaux et de la jurisprudence qui permet d'améliorer la conception des outils numériques.

***Merci Astrid, David, Winston
Telecom Paris,***

***Céline Barbosa- Avocat au Barreau de Paris
membre de l'ARCSI, membre du CEFYS***