

NOZHA BOUJEMAA



« Les algorithmes sont-ils transparents et éthiques ? Pour nous en assurer, nous avons besoin d'outils adaptés. »

Un algorithme peut-il être juste ?

Nozha Boujemaa : Pour répondre, un référentiel est indispensable. C'est précisément une des difficultés que l'on rencontre lorsqu'on s'intéresse à la qualité des algorithmes. La question à se poser est de savoir : « Pour qui ? ». La notion d'algorithme juste, ou loyal (d'une façon générale, on parlera de transparence), est différente selon que l'on se place du point de vue du consommateur du service adossé à l'algorithme ou du producteur de ce service.

Le consommateur individuel est un citoyen, et de fait les questions de transparence algorithmique ont souvent été soulevées par des affaires de libertés individuelles et de droits civiques. Quand l'utilisateur est une entreprise ou bien même les services de l'État, les problèmes possibles sont ceux, d'une part, de la concurrence déloyale, des abus de position dominante et, d'autre part, de souveraineté nationale.

BIO EXPRESS

17 JUILLET 1963
Naissance

2010-2015
directrice du centre
de recherche Inria
Saclay Ile-de-France.

2017
Directrice du projet
TransAlgo, de l'Inria
et de l'institut Dataia.

La question de la transparence est essentiellement celle de l'asymétrie informationnelle, entre celui qui propose un service logiciel et celui qui l'utilise, le second en sachant beaucoup moins que le premier. La réduction de cette asymétrie est cruciale : on doit savoir comment nos données sont calculées et par où elles transitent. Notons que la discrimination ou les biais potentiels ne sont pas toujours intentionnels.

Ces problématiques sont le cœur du projet *TransAlgo* que vous coordonnez. Comment est-il né ?

Nozha Boujemaa : Cette initiative fait suite à la loi pour une République numérique (promulguée le 7 octobre 2016) proposée par l'ancienne secrétaire d'État au numérique Axelle Lemaire, ainsi qu'au rapport commandé au Conseil général de l'économie (CGE) sur la gouvernance des algorithmes. La loi prévoit

notamment le droit à l'explication pour tout citoyen quant aux décisions algorithmiques prises par les services de l'État. C'est une avancée notable qui fait de la France un précurseur sur ces sujets aux niveaux européen et international.

En termes de protection des libertés individuelles, elle complète le Règlement général sur la protection des données (RGPD) adopté par le Parlement européen le 14 avril 2016. Ce dernier porte surtout sur les données personnelles, la portabilité, alors que la loi numérique s'intéresse à l'explication des décisions, celles-ci ne s'appuyant pas nécessairement sur des données personnelles. Le droit à l'explication est tout à fait original et anticipe bien l'évolution de la société où les algorithmes vont être de plus en plus amenés à prendre des décisions.

C'est une question de confiance dans l'État, et ses services, au moment où il entame sa transformation numérique. Il s'engage à cette occasion sur la transparence et l'explication des décisions algorithmiques. Il pourra alors répondre à des questions du type : « Pourquoi mon voisin et moi avons-nous été traités différemment alors que nous sommes dans la même situation ? »

Très tôt dans la genèse de la loi numérique, la nécessité de l'accompagner par un volet scientifique s'est imposée. En d'autres termes, on a pris conscience des difficultés techniques qui allaient se dresser devant le législateur. De fait, l'explicabilité des algorithmes (expliquer le cheminement de la prise de décision et sa traçabilité) n'est pas une question résolue, tant s'en faut. De plus, une des recommandations phare, du CGE était la création d'une plateforme scientifique dédiée à la régulation et à la gouvernance des algorithmes. C'est là qu'entre en scène Inria, sollicité pour aider à la montée en compétence des services de l'État en ce domaine.

Mais il a d'abord fallu se mettre d'accord sur les objectifs ?

Nozha Boujemaa : De notre point de vue, il est préférable de ne pas parler de gouvernance ni de régulation des algorithmes, parce que c'est un peu utopique... Ce n'est pas le bon objectif à se donner. Je crois que nous avons été compris, même si chez nos interlocuteurs la tentation a été grande, je peux vous l'assurer, d'imaginer une autorité des algorithmes, un système de certification de transparence, avec des labels...

Ce n'est pas notre rôle en tant qu'organisme de recherche : chacun son métier et la régulation n'est pas le nôtre. En

revanche, nous sommes compétents pour répondre à des questions difficiles sur l'explicabilité, l'auditabilité et la loyauté des algorithmes, sur les biais possibles à la fois des données et des algorithmes, les deux étant liés dans ce qu'il vaudrait mieux appeler un système algorithmique. Cette dualité est essentielle. C'est sur ces questions que se penche *TransAlgo*.

Il s'agit donc d'ouvrir la boîte noire à laquelle on compare souvent les algorithmes ?

Nozha Boujemaa : Nous réfléchissons à des outils dits d'auditabilité pour essayer de comprendre le fonctionnement d'un algorithme. Pour ce faire, on n'a pas nécessairement besoin du code source. L'une des approches est la rétro-ingénierie, ou *reverse engineering* : avec des jeux de données tests bien calibrés et dont on maîtrise les variations, on étudie le comportement d'un algorithme afin d'en comprendre le mieux possible les rouages. Par exemple, on peut étudier un algorithme en ne faisant varier que les informations concernant la géolocalisation. On voit alors dans quelle mesure elle influe sur les résultats de l'algorithme.

La rétro-ingénierie est particulièrement utile pour un type de boîte noire, les algorithmes dont le fonctionnement est bien connu du concepteur, mais que celui-ci ne communique pas.

On doit bien maîtriser le protocole d'expérimentation et la méthodologie. Par exemple, avec un algorithme de classement, tel un moteur de recherche, les résultats peuvent varier simplement parce qu'au cours du temps, le corpus sur lequel opère l'algorithme a changé. Les précautions qui s'imposent parfois rendent indispensable le travail interdisciplinaire : des mathématiciens, des informaticiens, mais aussi des juristes, des sociologues, des économistes collaborent sur ces questions-là et encadrent ensemble les expérimentations.

« Le droit à l'explication anticipe l'évolution d'une société où les algorithmes vont prendre plus de place »

L'autre type de boîte noire concerne les algorithmes dont même le concepteur ne maîtrise pas tout le fonctionnement. C'est le cas de ceux issus de l'apprentissage profond, tels les réseaux de neurones, où les paramètres sont ajustés automatiquement, à mesure des exemples, pour remplir la tâche assignée. On doit alors se concentrer sur l'explicabilité. Des travaux de recherche sont indispensables pour rendre explicable ce qui ne l'est pas aujourd'hui.

Insistons sur le fait que tous les algorithmes de l'intelligence artificielle ne sont pas inexplicables, seuls ceux d'une famille particulière, relevant du *deep learning*, le sont.

Quelles approches sont explorées ?

Nozha Boujemaa : Dans cet élan, les approches d'*accountability* sont très intéressantes. Il s'agit de l'idée de rendre compte du comportement des algorithmes, d'évaluer leur loyauté en se penchant par exemple sur les données qui les ont alimentés et de repérer celles qui conduisent à des discriminations.

Il ne faut pas les éliminer, mais les identifier, les prendre en compte et maîtriser le processus de décision de sorte qu'elles n'influent pas dessus. On parle d'*Equity by design* (« équité par construction »).

L'élaboration de nouvelles démarches d'apprentissage fait aussi partie de nos objectifs, afin que les questions éthiques et juridiques soient prises en compte dès le départ : c'est la *transparency by design*.

Les constructeurs d'algorithmes sont-ils prêts à vous suivre ?

Nozha Boujemaa : Au départ, plusieurs industriels n'ont pas très bien pris notre démarche. Pour eux, elle s'apparentait à un manque de confiance en leurs produits. Ils pensaient qu'on allait traquer les failles, les pointer du doigt. Mais nous ne sommes pas des gendarmes. Notre rôle est de mettre à disposition des outils logiciels au ➤