

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 ➤

service de tout le monde, citoyen, industriel, autorités de régulation...

Après un long travail de pédagogie, nous avons assisté à un changement des mentalités à notre égard. Les industriels ont compris qu'on leur proposait un moyen d'établir un lien de confiance avec le consommateur. Or la confiance incite à l'achat...

On peut faire le parallèle avec l'agro-alimentaire ou l'industrie pharmaceutique, deux domaines où l'importance de ce lien de confiance est reconnue depuis longtemps. D'ailleurs on m'a un jour fait remarquer que les médicaments nécessitent une autorisation de mise sur le marché, pourquoi n'en serait-il pas de même avec les algorithmes ?

Cette remarque est pertinente, car en effet, on met en circulation des algorithmes dont on sait peu de choses en fin de compte. C'est pour ça que l'État s'est posé la question de la régulation, et que nous avons proposé de travailler sur les outils de la transparence.

Pour certaines questions, des travaux de recherche sont encore nécessaires et peuvent réclamer du temps. Mais du chemin a été parcouru : des interlocuteurs hier méfiants ont compris l'intérêt de ces approches et proposent aujourd'hui de financer des thèses sur le sujet.

Comment faire admettre ce genre d'idée aux Gafam ?

Nozha Boujemaa : Il se trouve que ces Gafam (Google, Amazon, Facebook, Apple et Microsoft), et d'autres encore, sont à l'origine de Partnership on AI, une initiative visant à établir de bonnes pratiques dans les systèmes d'intelligence artificielle et à éduquer le public à ce sujet. Ils ont compris la vague citoyenne mondiale autour des questions d'éthique, de transparence, de discrimination, et ont pris les devants après la médiatisation de certains abus.

Il y a clairement une part de marketing, mais il y va aussi de leur survie commerciale. Aux États-Unis, de nombreux rapports et livres ont pointé du doigt les risques et certaines pratiques douteuses. Leur business menacé, ces grandes entreprises se devaient de réagir.

La clé de voûte de TransAlgo n'est-elle pas le volet pédagogique ?

Nozha Boujemaa : C'est en effet un point très important dans nos objectifs. Aujourd'hui, des groupes de travail thématiques ont commencé à réfléchir sur la définition des concepts que nous avons

évoqués (juste, loyal, éthique, responsable...) et à identifier ce qui est objectif, invariable. Cela pose la question de la relativité des concepts. Un site dédié est en cours de production, avec en page d'accueil cet aspect pédagogique essentiel.

Nous avons commencé à rédiger des textes synthétiques sur ces concepts et à sélectionner des cas d'usage emblématiques, par exemple sur la variabilité des prix, pour interpeller un large public. En complément, nous ouvrirons un espace de ressources avec des articles, des rapports, des vidéos, des logiciels en open source.

Et sur le plan scientifique ?

Nozha Boujemaa : En parallèle à ce volet pédagogique, nous lançons des groupes de travail scientifiques thématiques sur les algorithmes de classement, de recommandation, d'apprentissage, le contrôle de l'usage des données, les biais des données et des algorithmes. L'idée est de constituer une communauté scienti-

seront étudiées sans oublier leurs interfaces avec les sciences humaines et sociales. Une des priorités définies porte sur la transparence et l'éthique des systèmes algorithmiques. Le projet, prévu pour huit ans et demi et financé à hauteur de dix millions d'euros, rassemblera cent trente enseignants-chercheurs autour de quatorze partenaires académiques (Inria, CEA, École polytechnique, HEC, Ensae, Télécom Paris Tech, Centrale, Supélec, université Paris Sud, CNRS...).

Le numérique n'ayant pas de frontières, qu'en est-il de la dimension internationale ?

Nozha Boujemaa : Elle est fondamentale ! Nous ne pouvons pas nous contenter d'avoir notre propre définition de la transparence, de l'équité... Si c'était le cas, ce serait un coup d'épée dans l'eau. Aussi, dès le début du projet, nous avons discuté avec nos homologues du Data Science Institute, à l'université Columbia, aux États-Unis,

« Les Gafam ont compris la vague citoyenne mondiale portant sur les questions d'éthique, de transparence... »

fique de recherche sur le sujet réunissant plusieurs disciplines scientifiques (les télécommunications, la gestion de bases de données, l'apprentissage...), mais aussi des juristes, des économistes, des sociologues... Ce que produira cette plateforme scientifique sera ouvert à tous, vérifiable, testable.

Plusieurs autorités de régulation sont aussi intéressées pour se joindre à cette plateforme et l'accompagner. Le Conseil national du numérique (CNNum) est également très impliqué. Le secrétariat d'État au numérique suit *TransAlgo* de près.

Mentionnons également Dataia, l'institut interdisciplinaire en sciences de données, intelligence et société que nous lançons (un des Instituts convergences financés par l'ANR et les Investissements d'avenir). Les sciences des données y

avec la National Science Foundation, aux États-Unis, et avec le programme Big Data Crest, de la Japan Science and Technology agency, au Japon, ainsi qu'avec l'institut Alan Turing, en Angleterre. L'idée de transparence algorithmique fait son chemin, malgré les différences marquées des sphères culturelles. Avec mes collègues américains, nous travaillons à ce que ce domaine bénéficie de crédits de recherche.

Nous avons aussi noué des contacts avec l'agence qui conseille le premier ministre italien sur l'agenda numérique, avec leurs homologues auprès du ministère du numérique espagnol... Il importe d'avancer ensemble. La mise en route est difficile, mais un mouvement international est indéniablement lancé. ■

PROPOS RECUEILLIS PAR LOÏC MANGIN