La transparence des algorithmes face à l'Open Data : Quel statut pour les données d'apprentissage ?

Danièle Bourcier, Primavera de Filippi

La transparence des algorithmes face à l'Open Data

Quel statut pour les données d'apprentissage ?

Danièle Bourcier
Directrice de recherche CNRS

Primavera de Filippi
Chargée de recherche CNRS

CERSA CNRS/Université Paris 2

Mots clés: open data ; accès aux données intermédiaires, données privées, réseaux de neurones artificiels, algorithmes, apprentissage machine

Présentation

A l'heure du numérique, caractérisé par la collecte et l'utilisation généralisée de données sur le monde, nous avons besoin d'une administration de plus en plus ouverte.[1] Le secteur privé a été le premier à s'emparer des nouvelles opportunités offertes par ces grands jeux de données (*big data*) et les nouvelles technologies de gestion de données (*data management)* : gageons que le gouvernement utilise à son tour ces techniques afin d'optimiser et d'automatiser ses activités, y compris pour les décisions administratives.[2] Mais à la différence du secteur privé, le secteur public a des exigences et des contraintes supplémentaires vis-à-vis du public, dont il doit tenir compte.

En effet, en dépit de leur efficacité, une des premières craintes soulevées par l'utilisation d'algorithmes informatiques au sein des administrations publiques est

[1] D. Bourcier & P. de Filippi, "Présentation," in *Open data, Big data, nouveaux défis pour la vie privée,* D. Bourcier & P. de Filippi (eds), Mare & Martin, 2016 p. 23

[2] Bourcier, D., De Filippi, P. (2018). « Les algorithmes sont ils devenus le langage ordinaire de l'administration ? » in Cluzel, L., Koubi, G., Tamzini, W. (eds) Lectures critiques du Code des relations Public et administration. Editions Lextenso. pp. 193-210

le risque d'*obfusquer* les fondements factuels et juridiques des décisions administratives.[3]

Le *big data* se réfère à la collecte et l'agrégation de grandes masses de données dans le but d'extraire de nouvelles informations (ou données). En effet, de plus en plus les données sont imbriquées dans des chaînes de traitement avant d'être utilisées pour une finalité particulière (science des données). C'est le cas notamment des réseaux de neurones ou de l'apprentissage machine (*machine learning*). Il ne s'agit plus seulement d'encadrer le recueil de données à la source (auprès de l'utilisateur par exemple) mais aussi de réguler les opérations d'apprentissage des algorithmes qui font appel à des technologies de plus en plus complexes pour traiter ces données.

Face à une algorithmisation croissante de l'administration, le gouvernement doit désormais construire un droit à la transparence qui soit concrètement applicable (Brin, 1998). Cela s'est traduit premièrement par les politiques d'ouverture des données publiques (*open data*), et ensuite, depuis 2016, par les exigences de transparence des décisions fondées sur des algorithmes.[4]

Le régime actuel de l'open data est-il suffisant à encadrer ces nouvelles façons d'alimenter les algorithmes ? La loi serait-elle déjà en retard sur les précautions à prendre face aux technologies de l'apprentissage machine ou des réseaux de neurones ?

Cet article s'intéresse aux problématiques liées à l'application d'algorithmes dans les décisions administratives, et plus particulièrement aux données et méthodes utilisées dans ces applications.[5] Le législateur s'est jusqu'à présent focalisé sur la transparence de la "décision" qui n'est qu'un certain type de traitement de données. Mais rien n'est précisé sur les données qui vont influencer ces algorithmes, c'est à dire les données qui se trouvent *en amont de la décision*. On fait l'hypothèse dans

[3] De Filippi, P. (2017). "Repenser le droit à l'ère numérique : entre la régulation technique et la gouvernance algorithmique", in Gautrais, V, Moyse P.E. (eds.) Droit et Machine.Vol.3 Éditions Thémis. pp-53-96

[4] Loi pour une République numérique N° 2016-1321 du 7 octobre 2016 (Loi Lemaire).

[5] On a beaucoup discuté d'un nouveau "droit à l'explication" (*right to explanation*) qui émergerait de ces algorithmes dans le cadre de l'article 22 du Règlement général sur la Protection des données (RGPD).

cet article que ce droit à l'explication devrait porter aussi sur les données utilisées pour entraîner ces algorithmes.

On fera d'abord un rappel de l'évolution des politiques d'Open data[6], puis on parlera des nouvelles tendances vers l'algorithmisation du droit et de l'administration dans le contexte du gouvernement ouvert et le rôle joué par les données au sein de ces nouveaux processus décisionnels. [7] Enfin, on analysera la difficulté d'assurer une réelle transparence pour de nouveaux types d'algorithmes (e.g. les algorithmes d'apprentissage automatique) qui seront de plus en plus utilisés au sein de l'administration. Nous soulignerons notamment la nécessité - actuellement encore peu explorée - de garantir non seulement l'accès au code source de ces algorithmes, mais aussi l'accès aux bases de données qui les ont entraînés, ainsi qu'aux critères de sélection utilisés pour construire ces base d'apprentissage.

1- De l'accès aux documents à l'ouverture des données

Le principe d'ouverture des données publiques a été posé en 1978 en France à travers le ***principe d'accessibilité*** aux dossiers de l'administration et par opposition à la tradition du secret et de l'opacité qui régnait dans le secteur public. La loi du 17 juillet 1978 [8] fut d'abord conçue comme consacrant un droit d'accès aux documents administratifs par les administrés. La notion de document communicable était soumise à des limitations et exceptions (notamment sous le couvert de la protection des données personnelles ou du secret d'état) et surtout elle supposait que l'administré fasse une démarche volontaire auprès des services concernés. Dans cette conception, la publication n'est qu'une possibilité offerte aux administrations.

[6] Bourcier, B., De Filippi, P. (2013), "**L**'Open Data : universalité du principe et diversité des expériences ?", dans *La Semaine Juridique*, JCP Lexis-Nexis, Septembre 2013; De Filippi P., Bourcier, D. (2014), « Open Data » : l'ouverture des données in *La Semaine Juridique Administrations et collectivités territoriales (JCPA, Lexis Nexis)*, n.28, Janvier 2014

[7] Est ce que le principe de transparence des algorithmes peut être appliqué dans des situations ou on ne va pas rendre accessible les données qui iront alimenter ces algorithmes (e.g. dans le cas des données non-publiques, car collectées ou générés par des acteurs privés). De plus en plus l'administration publique va s'appuyer sur des données privées; peut-on obliger les acteurs privés à publier leurs données?

[8] Modifiée par la loi du 12 avril 2000 et l'ordonnance du 6 juin 2005

Changement de perspective: l'Ordonnance du 6 juin 2005, en transposant la directive du 17 novembre 2003, s'est alignée sur le droit européen. Il ne s'agissait plus de documents administratifs mais d'information publique[9]. La particularité de cette loi résidait dans le fait qu'au-delà de l'ouverture des données, elle instaurait le droit de les ***réutiliser***.

Cependant la réutilisation impose que ces données ne peuvent être dénaturées ou altérées dans leur signification, et que les sources et la date de leur dernière mise à jour soient mentionnées.[10] Ces conditions figureront dans la Licence Ouverte (LO) mise en place par le gouvernement français. Cette licence, mis à part ces conditions restrictives dues à la nature officielle des données, correspond à la licence Creative Commons BY[11], qui figure d'ailleurs parmi la liste de licences compatibles mais devant cependant faire l'objet d'une "homologation"[12] justifiée. En effet seules deux licences sont autorisées par le décret[13] d'application pour les administrations publiques: la licence Ouverte d'Etalab et l'Open Database License (ODbL). Les autres licences libres de réutilisation doivent faire l'objet d'une procédure d'homologation. Les licences Creative Commons, internationales et traduites dans plus de 20 langues devraient cependant pouvoir être autorisées dès maintenant - à la demande de certaines administrations de santé notamment- car la nouvelle licence 4.0 a intégré le régime du droit *sui generis* des bases de données, comme ODbL. Enfin, un portail interministériel français unique a été mis en place par la mission Etalab placé sous l'autorité du Premier ministre (décret du 21 février 2011).

---

[9] J. Chevallier, "Le droit français et la question des données publiques", in D. Bourcier & P. de Filippi (eds), 2016, op. cit. p. 33.

[10] La loi de 1978 (Article 12) déclare, par principe, la libre réutilisation des données. Toutefois, sauf accord préalable de l'autorité concernée, les réutilisations : (1) ne doivent pas altérer les informations publiques, c'est-à-dire changer leur état, leur signification, (2) ne doivent pas dénaturer leur sens, c'est-à-dire modifier leur nature ou leur qualité, (3) doivent mentionner « leurs sources et la date de leur dernière mise à jour », c'est-à-dire indiquer les noms des administrations qui les ont produites et la date de leur production ou mise à jour ;

[11] www.creativecommons.fr

[12] La licence **Creative Commons Attribution - Partage dans les mêmes conditions (CC-BY-SA) 4.0** a finalement été homologuée par la décision d'homologation du 26 septembre 2017 pour le périmètre des levés bathymétriques, dalles bathymétriques, natures de fonds, délimitations maritimes, câbles et conduites sous-marines, toponymes, et épaves et obstructions du Service hydrographique et océanographique de la marine (SHOM) jusqu'au 25 septembre 2020. Mais cette homologation n'est pas générale : elle est liée à une demande précise. Dans ce cas, celle du SHOM, qui a dû justifier, compte tenu de ses missions militaires, de la nécessité d'une licence internationale...

[13] Décret n° 2017-638 du 27 avril 2017 relatif aux licences de réutilisation à titre gratuit des informations publiques et aux modalités de leur homologation.

Enfin, la loi Lemaire (2016) a précisé un certain nombre de dispositions sur les données publiques, et notamment sur la notion de données de référence,[14] et introduit de nouvelles obligations en ce qui concerne la transparence des algorithmes.

Les textes législatifs et réglementaires décrivant les conditions que les opérateurs responsables du traitement des données doivent remplir portent essentiellement sur les données personnelles. Par définition, elles ne font pas partie de l'Open data.[15] En revanche, la loi Lemaire (Code des Relations entre le public et l'administration, art. 321-7 et art. 321-8) évoque les principales conditions auxquelles les données de référence doivent se soumettre : fiabilité, disponibilité, sécurité, ainsi que le "maintien en conditions opérationnelles" et "performance de mise à disposition". Mais ces conditions restent vagues sur la façon dont les données doivent être gérées.[16]

D'autre part, la loi Lemaire introduit aussi la possibilité pour les administrations de communiquer aux citoyens « *les règles* » et « *principales caractéristiques* » de mise en œuvre des traitements algorithmiques servant à prendre des décisions individuelles les concernant. Il s'agit donc là de la communication des règles et non pas des données.

Si l'on intègre ces différents textes, on peut extraire une liste d'information qui feraient l'objet de l'Open data :

-Les *documents* communiqués sur demande par la Commission d'accès aux documents administratifs (CADA);

-Les *documents* qui figurent dans le répertoire des principaux documents administratifs (que doivent tenir les administrations).

-Les « *bases de données publiques* », dont les bases de données juridiques comme LEGIFRANCE, complétées ou reconfigurées;

[14] Le nouveau Service public de la donnée en effet vise à mettre à disposition, en vue de faciliter leur réutilisation, des jeux de données de référence, c'est-à-dire ceux qui présentent le plus fort impact économique et social, pour que les entreprises puissent les réutiliser pour leurs services. À ce jour, neuf jeux de données sont inclus dans ce service (notamment, SIRENE, base adresses, diverses bases de données géographiques etc.).

[15] De Filippi, P., Maurel, L. (2014). The Paradoxes of Open Data and how to get rid of it : analysing the interplay between Open Data and sui-generis rights on databases, in *International Journal of Law and Information Technology*, Oxford University Press. 2014, 0, 1-22.

[16] De Filippi, P., Bourcier, D. (2013), " La double face de l'Open data", dans *Petites Affiches*, Octobre 2013

-Les *données de référence*, c'est-à-dire les données « *dont la publication présente un intérêt économique, social, sanitaire ou environnemental* »;

-Les *règles qui régissent les algorithmes* utilisés par les administrations publiques pour prendre des décisions.

Précisons que la loi Lemaire prévoit l'ouverture des données de jurisprudence : mais le processus de mise en œuvre, lancé dans le cadre d'un groupe de travail réunissant les parties prenantes (services judiciaires, Legifrance, Cour de cassation, Conseil constitutionnel, Conseil d'Etat, CNIL) n'a donné lieu pour l'instant à aucun texte d'application. Le fonds documentaire (jurisprudence judiciaire) de la base de données LEGIFRANCE comprend les arrêts, publiés ou non, de la Cour de Cassation mais au niveau des Cours d'appel, il n'est question que d'une "sélection" d'arrêts sans en préciser les critères. Enfin, il n'existe pas de corpus systématique des jugements des Tribunaux de première instance[17]. Cette disposition intéresse pourtant le secteur du *legaltech* qui développe leurs algorithmes prédictifs à partir des données jurisprudentielles.[18]
En revanche, le "code" des algorithmes fait partie des documents communicables sauf s'ils menacent la sécurité publique, le secret des affaires et les droits de propriété intellectuelle. Pour la CADA, les codes sources des logiciels de calcul de la taxe d'habitation et de la taxe foncière, par exemple, ont vocation à être rendus publics par l'administration fiscale.[19]

[17] D'après le site Legifrance, ce fonds jurisprudentiel comprend :
les grands arrêts de la jurisprudence civile en texte intégral ;
les décisions de la Cour de cassation :
publiées au *Bulletin des chambres civiles* depuis 1960,
publiées au *Bulletin de la chambre criminelle* depuis 1963,
ainsi que l'intégralité des décisions, publiées ou non, postérieures à 1987.
des décisions des cours d'appel et des juridictions de premier degré ;
une sélection de décisions du Tribunal des conflits publiées au Bulletin depuis 1993.

[18] Voir, dans section 2-A, les algorithmes en matière judiciaire, dont l'apprentissage se fait sur une sélection d'arrêts.

[19] Avis CADA n° 201445 du 8/01/2015 : La commission considère que "l'appréciation de l'administration selon laquelle la réutilisation envisagée se heurterait à des difficultés techniques, voire à une impossibilité matérielle, ne saurait fonder le refus de communiquer le document sollicité dans l'état où l'administration le détient".
La commission a donc émis "un avis favorable à la communication à Monsieur X du code source sollicité, sous la forme sous laquelle l'administration le détient. Le demandeur est libre de le réutiliser dans les conditions fixées à l'article 12 de la loi du 17 juillet 1978, en l'absence de droits de propriété intellectuelle détenus par des tiers à l'administration, dont le directeur général des finances publiques ne fait pas état".

2- L'algorithmisation de l'administration publique : Applications et réponses législatives

Les systèmes experts opéraient à partir d'un moteur d'inférence (logiciel) et d'une base de connaissances qui constituait le coeur de l'expertise. Cette base était d'ailleurs soumise au droit d'auteur. Les systèmes de *machine learning* prolongent les systèmes experts mais en diffèrent au moins sur deux points: *"d'abord ils n'ont pas besoin d'être programmés mais peuvent apprendre à partir d'exemples, et ensuite, parce qu'ils utilisent la théorie des probabilités, ils représentent mieux les paramètres du monde réel avec le bruit concomitant, les exceptions, les ambiguités, et l'incertitude des résultats."* [20] La "base de connaissances" est une base d'apprentissage constituée de données, souvent hétérogènes et dynamiques: ce qui signifie que les données d'apprentissage au temps *t* (y compris leur périmètre et leur volume, qui sont des éléments fondamentaux pour évaluer la pertinence et la fiabilité du système) ne sont pas *a priori* obligées d'être "ouvertes". Rien ne contraint les concepteurs à décrire leurs méthodes de conception de ces bases, alors que la question de l'ouverture des algorithmes, nécessairement plus obscurs, est devenue paraît-il cruciale pour l'intelligibilité de la décision.

A. Une typologie des algorithmes dans l'administration publique

L'intelligence artificielle avait donné lieu à de nombreux travaux académiques à la fin des années 1990. Les applications dans l'administration étaient nombreuses mais limitées en taille et en portabilité. En droit, un courant scientifique "Intelligence artificielle et Droit" avait développé une riche réflexion cognitive (plus que technique) autour de la question "Comment raisonnent les juristes?". Depuis quelques années avec le Big Data, un courant pragmatique dit *LegalTech* tente de développer des applications juridiques pour les professions judiciaires. La portée de cette "révolution" diffère suivant les finalités.

[20] E.Alpaydin, *Machine learning, The new AI,* The MIT Press, Essential knowledge series, 2016 p.52 (notre traduction).

Il convient d'abord de distinguer les ***opérations de gestion*** et les ***systèmes de décision***. Certaines applications sont des aides aux professions judiciaires (prédictibilité statistique), mais restent des applications non opposables aux citoyens, alors que d'autres systèmes visent à faciliter voire automatiser les décisions publiques (justice, administration) susceptibles de faire grief.

Appartiennent à cette première catégorie (*opérations de gestion*) les quatres systèmes suivants :

L'Internal Revenue Service (IRS), agence rattachée au département du Trésor des Etats-Unis utilisent divers programmes lui permettant de détecter d'éventuelles lacunes dans les déclarations fiscales des citoyens. Ses agents ont d'abord développé un programme de contrôles aléatoires : le **Taxpayer Compliance Measurement Programme** (TCMP), méthode statistique qui attribue des notes à partir de contrôles effectués sur un échantillon de déclarations de revenus. Cette analyse discriminante permet d'identifier les déclarations de revenus des individus et des entreprises réfractaires au contrôle fiscal. Ce type de décision est une décision interne qui ne prive les usagers d'aucun droits. Mais les modes d'échantillonnage restent très opaques.

De même, en 2014, le gouvernement français avait souhaité, à l'image des autres Etats, faire usage d'algorithmes dans un but de lutte contre la fraude fiscale. Avec **OpenFisca**,[21] un algorithme avait donc été mis en place par la Direction générale des finances publiques à titre expérimental. Ensuite, par un arrêté publié le 14 novembre 2017[22] Bercy avait officialisé les traitements automatisés de lutte contre la fraude fiscale en raison des chiffres importants d'omission fiscale. L'arrêté prévoyait que «*les traitements mis en œuvre peuvent utiliser, d'une part, à titre pérenne les données des professionnels et des personnes physiques en lien avec une entreprise et, d'autre part, à titre expérimental les données des particuliers sans lien avec une entreprise, contenues dans la base*». Précaution utile; il a été finalement décidé que les données utilisées seraient fixées par le législateur.

[21] OpenFisca modélise le système socio-fiscal français en code informatique. Pour plus d'informatoin, voir le site Internet https://fr.openfisca.org/ Il prévoi,t selon le s termes du site, et dans la continuité de beaucoup de travaux de légistique, de 'transformer la loi en logiciel'.

[22] Arrêté du 28 août 2017 modifiant l'arrêté du 21 février 2014 portant création par la direction générale des finances publiques d'un traitement automatisé de lutte contre la fraude dénommé « ciblage de fraude et valorisation des requêtes »

Encore, en France, la société **Predictice**[23] a pour ambition de prévoir les probabilités du résultat d'une procédure contentieuse par le biais d'un algorithme. Cet algorithme serait capable de prévoir les chances de succès ou d'échec d'une procédure, le montant des indemnités que l'on a des chances d'obtenir, et d'indiquer quelle juridiction est susceptible de prendre la décision la plus favorable aux intérêts du client. Pour opérer, cet algorithme s'appuie sur le Big data, une base de données de 1,5 millions de décisions de justice. On ne connaît pas plus de détails sur la pertinence de ce volume et des choix opérés.

Un autre site de justice prédictive (en Open source) s'est spécialisé en droit administratif français: le site **Supra Legem,** désormais racheté par les éditions Lefèbre Sarrut. L'algorithme reprend la jurisprudence administrative disponible sur Légifrance et propose là aussi des outils d'analyses et de recherches fondés sur les traitements algorithmiques. Ces statistiques ont pour objectif de fournir l'issue d'un litige devant une juridiction administrative[24]. Là aussi, les données qui ont permis cet apprentissage seraient utiles pour évaluer la pertinence des résultats.

Différents de ces outils de gestion, d'autres systèmes sont utilisés pour *simuler des décisions*. Par exemple, certains Etats des Etats-Unis utilisent des outils algorithmiques dans le domaine de la justice qui déterminent directement la prise de décision. On peut citer à titre d'exemple le logiciel **COMPAS** (Correctional Offender Management Profiling for Alternative Sanctions) : un algorithme prédictif utilisé depuis 1998 par les Etats de New York, de Californie, de Floride et du Wisconsin, pour faciliter la gestion des dossiers des détenus. Alors que ce logiciel avait comme but d'assister les administrations dans la préparation des procès, le choix des modalités de détentions, et l'identification des meilleurs programmes de réinsertions, il est aujourd'hui utilisé principalement par les juges pour statuer sur la liberté conditionnelle. Cet algorithme s'appuie sur plus d'une centaine de critères pour estimer le risque de récidive d'une personne condamnée

[23] https://www.dalloz-actualite.fr/interview/l-utilisation-de-l-outil-predictice-decoit-cour-d-appel-de-rennes
Le logiciel testé par des magistrats a été abondamment critiqué : "*Le logiciel ne s'intéresse qu'au dispositif d'une décision de justice. L'algorithme ne sait pas lire toutes les subtilités de la motivation, surtout lorsque la décision est complexe."*

[24] Cela serait impossible en droit français, pour l'instant.

pénalement. Or, bien que certains critères aient été communiqués au public,[25] la plupart de ces critères restent inconnus car considérés comme des informations confidentielles relevant du secret commercial de l'éditeur.

De même au Royaume-Uni, la police de la ville de Durham s'est dotée d'un programme similaire nommé « Harm Assessment Risk Tool » afin d'aider les policiers à décider du placement ou non d'un suspect en détention provisoire. Ce logiciel repose sur des archives de la police de la ville récoltées entre 2008 et 2012. Les données comprennent le casier judiciaire de la personne, son âge, le type d'infraction commise et son code postal. L'algorithme doit analyser les décisions prises par les agents pendant toute cette période et observer la récidive ou non des suspects afin d'évaluer le risque de récidive[26]. La question des faux positifs et des faux négatifs n'est jamais évoquée dans l'élaboration de la base d'apprentissage.[27]

B. Premières tentatives de régulation des algorithmes

Face à ce déferlement d'annonces de décisions algorithmiques, des lois ou projets de loi fleurissent dans tous les Etats. En France, les premiers textes faisant référence à la décision automatique datent de la loi Informatique et Libertés de 1978[28]: l'article 2 distinguait la décision de justice, la décision administrative et la décision privée. Comme une forme de "gradation" dans la prise de décision, la décision de justice ne pouvait, par principe, être "automatisée": *"Aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement automatisé de données* ***à caractère personnel*** *destiné à évaluer certains aspects de sa personnalité."*

Cet article a été modifié *par la loi n°2004-801 du 6 août 2004 :* Le premier alinéa sur la décision de justice a été gardé tel quel mais le deuxième alinéa a supprimé la référence à la "décision administrative et privée" pour devenir :"'*Aucune autre*

[25] Parmi ces critères, on trouve les autres affaires en cours dans laquelle la personne est impliquée; son casier judiciaire; la stabilité de son logement, la stabilité de son emploi, l'éventuel usage de drogue par celle-ci, son intégration sociale, etc.

[26] Une expérimentation de ce programme mise en place en 2013 pour deux ans, avait révélé que les prévisions de risque faible de récidives étaient justes à 98%. Les prévisions de risque élevé l'étaient à 88%. Cela pose la question de la marge d'erreur inhérente à ce genre d'algorithme, et aux risques de suivre automatiquement l'aide apportée par ce logiciel prédictif.

[27] E. Alpaydin, op.cit. P. .53 : le faux négatif peut avoir une impact plus crucial que le faux positif dans le cas de décision faisant grief.

[28] Déjà citée

décision produisant des effets juridiques à l'égard d'une personne ne peut être prise sur le ***seul fondement*** *d'un traitement automatisé de données destiné à définir le profil de l'intéressé ou à évaluer certains aspects de sa personnalité."*

La discussion a donc longtemps portée sur le fait que l'intervention humaine, celle du décideur comme le juge, ou l'autorité administrative suffisait à interrompre le caractère automatique du résultat et de son opposabilité à l'usager. Désormais, on considère aussi que par défaut l'usager impliqué dans la décision (c'est-à-dire ayant pu poser des questions sur les fondements de la décision) lève aussi le caractère automatique de la décision: *"Ne sont pas regardées comme prises sur le seul fondement d'un traitement automatisé les décisions prises dans le cadre de la conclusion ou de l'exécution d'un contrat et pour lesquelles la personne concernée a été mise à même de présenter ses observations, ni celles satisfaisant les demandes de la personne concernée."*

Ces dispositions, dont le statut juridique n'était pas clair en termes d'obligations et de sanctions,[29] ont été précisées par l'article 22 du Règlement général (UE) 2016/679 du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel qui spécifie que "*les personnes ont le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé produisant des effets juridiques les concernant ou les affectant de manière significativ*e" et que "*les responsables de traitements doivent informer expressément les personnes que le traitement mis en œuvre entraîne une prise de décision automatisée ou un profilage*" (articles 13 paragraphe 2, sous f), et article 14 paragraphe 2, sous g)).

Ces règles se sont étendues progressivement avec l'adoption du principe de transparence[30] dans la décision individuelle. La loi du 7 octobre 2016 pour une

[29] De Filippi, P. (2016), "Gouvernance algorithmique : vie privée et autonomie individuelle à l'ère des Big data", in Bourcier, D., De Filippi, P. (eds.) Open Data & Data Protection : Nouveaux défis pour la vie privée. Mare & Martin

[30] D'autres textes normatifs sur la décision publique dans différents secteurs ont lancé cette notion de transparence mais il s'agit de transparence des informations. Ainsi en matière nucléaire, la loi n° 2006-686 du 13 juin 2006 relative à la transparence et à la sécurité en matière nucléaire art 1 (Abrogé en 2012) donne incidemment une définition de la transparence comme "l'ensemble des *dispositions prises pour garantir le droit du public* à une information fiable et accessible en matière de sécurité nucléaire". En matière de santé, conformément à l'article L. 1453-1 du code de la santé publique, on demande désormais aux entreprises commercialisant des produits sanitaires de rendre publics les avantages, les rémunérations accordés aux acteurs ainsi que l'existence des conventions conclues avec ces acteurs. Il existe même une base de données publique *Transparence - Santé* qui rend accessible les informations déclarées par les entreprises sur les liens d'intérêts qu'elles entretiennent avec les acteurs du secteur de la santé.

République numérique, dite loi Lemaire, a créé un principe intermédiaire de communication des algorithmes. Le décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique précise les modalités de la demande et de la communication des règles définissant un traitement algorithmique lorsque celui-ci a participé au fondement d'une décision individuelle. A l'article L. 311-3-1 du code des relations entre le public et l'administration, sont insérés désormais les articles R. 311-3-1-1 et R. 311-3-1-2 ainsi rédigés :

> *« Art. R. 311-3-1-1.-La mention explicite prévue à l'article L. 311-3-1 indique la finalité poursuivie par le traitement algorithmique. Elle rappelle le droit, garanti par cet article, d'obtenir la communication des règles définissant ce traitement et des principales caractéristiques de sa mise en œuvre, ainsi que les modalités d'exercice de ce droit à communication et de saisine, le cas échéant, de la commission d'accès aux documents administratifs, définies par le présent livre.*

Le principe de transparence en droit est donc lié au contrôle des algorithmes et non des données. Déjà en 1992, le Conseil constitutionnel avait reconnu pour la première fois l'objectif de valeur constitutionnelle d'accessibilité et d'intelligibilité de la loi.[31] *L'accessibilité* se rapporte à la possibilité de trouver physiquement le droit applicable tandis que *l'intelligibilité* renvoie à la lisibilité du texte et à l'adoption de dispositions suffisamment précises et de formules non équivoques.

Dans le monde des algorithmes, l'accessibilité exige que l'on puisse accéder au code source des algorithmes décisionnels, alors que l'intelligibilité implique la nécessité de "comprendre" le processus de délibération qui a porté à une décision donné. Bien que pour certaines catégories d'algorithmes le plus "intelligibles" (tels que les systèmes experts), l'accès au code source permettra de comprendre le fonctionnement et donc le raisonnement de ces algorithmes, pour d'autres catégories (tels que les systèmes d'apprentissage automatique) l'intelligibilité ne peut être aussi facilement assurée (Pasquale, 2015).
Le régime du droit à la transparence pour la décision algorithmique a été défini plus clairement dans la loi Lemaire: doivent désormais être précisés dans le

[31] Cet objectif a été élaboré d'abord par une décision du 16 décembre 1992, et précisé par une décision du 27 juillet 2006.

processus de décision, les données et la source des données utilisées par l'algorithme, les paramètres de traitement et les opérations effectuées, ainsi que le degré et le mode de contribution du traitement algorithmique à la prise de décision.[32]

Mais de quelles données s'agit-il? La diffusion des sources implique-t- elle les méthodes pour y accéder? Aucun des textes sur la transparence des algorithmes n'a pris en considération explicitement les données sur lesquelles ces algorithmes ont été entraînés.

3. Les données d'apprentissage relèvent-elles de l'Open data?

Tant au niveau de la description des applications que des régulations actuelles sur l'ouverture des données, la question des données d'apprentissage n'a été évoquée. Or, dans le cas des algorithmes d'apprentissage automatique, l'exigence de transparence ne peut pas se limiter à la mise à disposition du code informatique qui sous-tend ces algorithmes. La transparence des algorithmes nécessite aussi une plus grande clarté en ce qui concerne les critères de sélection des données d'entraînement et la définition du périmètre de ces données. De plus, dès lors que ces algorithmes sont entraînés sur des jeux de données de statuts différents, y compris des données privées, il convient de réfléchir à la possibilité (ou au besoin) d'exiger une mise à disposition de ces données, afin de pouvoir vérifier du bien fondé des décisions qui en sont issues.

A. Transparence au niveau des critères de sélection

[32] « Art. R. 311-3-1-2.-L'administration communique à la personne faisant l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, à la demande de celle-ci, sous une forme intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi, les informations suivantes :

« 1° Le degré et le mode de contribution du traitement algorithmique à la prise de décision ;

« 2° Les données traitées et leurs sources ;

« 3° Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé ;

« 4° Les opérations effectuées par le traitement ».

La transparence des algorithmes permettrait donc de remplir les objectifs d'un gouvernement ouvert, où les citoyens sont mis au courant des décisions prises par les administrations publiques, ainsi que des motivations sous-jacentes à ces décisions. Cependant, il est important de remarquer que tous les algorithmes n'opèrent pas de la même façon. Alors qu'il suffit de regarder leur code source afin de comprendre le fonctionnement de certains algorithmes (*e.g.* les systèmes experts, fondés sur des règles précises et prédéfinies à l'avance par des humains), d'autres algorithmes présentent des caractéristiques particulières qui en rendent impossible la compréhension par une simple observation du code. C'est le cas notamment des réseaux de neurones (*neural networks*) et des algorithmes d'apprentissage automatique (*machine learning*) qui s'appuient sur des grands jeux de données afin de développer leur propre réseau de neurones et leur propre système de raisonnement - deux éléments déterminants pour toute prise de décision algorithmique.

Ainsi, à la différence des systèmes experts, qui ne font que traiter les données qui leur sont fournies en entrée, les algorithmes d'apprentissage automatique s'alimentent, eux, d'un flux de données, qui contribuent - à chaque itération - à en faire évoluer leur mécanisme de raisonnement. Il existe notamment deux catégories de données qui vont influencer les décisions de cette dernière typologie d'algorithmes. D'une part, il y a les données qui sont envoyées à l'algorithme : les paramètres en entrée qui seront traités et analysés afin de déterminer, au cas par cas, le cheminement de l'algorithme dans son processus décisionnel. Il s'agit, par exemple, des données relatives aux individus concernés, telles que leurs caractéristiques démographiques, leurs occupations, leurs habitudes comportementales, ou encore leur casier judiciaire. D'autre part, il y a les données qui ont été utilisées préalablement pour entraîner l'algorithme, et lui permettre ainsi de développer une base de connaissance sur laquelle il fondera son processus de raisonnement.

L'utilisation croissante de ces algorithmes au sein de l'administration publique oblige à nous concentrer sur la nature et les caractéristiques des bases de données utilisées pour entraîner ces algorithmes. Puisque les réseaux de neurones des algorithmes de décisions se construisent à partir d'une base d'apprentissage

donnée, il n'est pas possible d'en analyser le fonctionnement en ne se concentrant que sur le code informatique qui régit ces algorithmes. Pour juger du bien fondé d'une décision algorithmique, il est nécessaire de prendre en compte non seulement l'exactitude des données en entrée, mais aussi l*'exhaustivité* et la *représentativité* des données sur lesquelles l'algorithme s'est entraîné, car celles-ci vont influencer le processus de raisonnement sur lequel la décision sera fondée.

Cela implique aussi de prendre en compte la manière dont ces données ont été collectées[33] : Est-ce qu'elles ont été échantillonnées de manière systématique ou aléatoire? La base d'apprentissage est-elle fiable et exhaustive? Et qui a présidé à la définition de cette base? Ces questions sont importantes car le raisonnement d'un algorithme d'apprentissage automatique va dépendre de la base d'apprentissage sur laquelle il a été entraîné. Ainsi, une même requête soumise à des algorithmes du même type, mais entraînés sur différents échantillons d'une même base de connaissances, fournira vraisemblablement des réponses différentes. La question de l'exhaustivité des fonds documentaires s'était déjà posée dans le contexte des requêtes posées aux bases de données juridiques (LEGIFRANCE par exemple). L'interrogation de ces bases de données par un juge (ou par un avocat) ne peut être pertinente que si le juge est au courant des critères de sélection et d'inscription des décisions juridiques dans la base de données : est-ce que la base de données comprend la totalité des décisions émises par toutes les cours d'appel, ou s'agit-il uniquement des décisions prises au sein d'une juridiction donnée? Ou simplement les décisions publiées donc déjà sélectionnées, mais par qui? La transparence des critères de sélection est fondamentale car si le juge se fonde sur une base de données incomplète ou biaisée, les résultats de la requête seront faussés et la décision qu'il était en train d'élaborer sur la jurisprudence applicable à son cas, remise en cause.[34]

De même, indépendamment de la façon dont il a été codifié, toute décision prise par un algorithme développé à partir d'une base de connaissances incomplète, non

[33] Bourcier, D., De Filippi, P. (2012), "Vers un nouveau modèle de partage entre l'administration et les communautés numériques", Nicolas Matyjasik, Philippe Mazuel (eds.) *Génération Y et gestion publique : quels enjeux* ?, Institut de la gestion publique et du développement économique (IGPDE)

[34] Une expérience avait été menée, avec les mêmes mots-clés, sur trois bases de données juridiques qui affichaient les mêmes contenus. Les réponses variaient d'un fonds à l'autre : voir D. Bourcier, "Le droit médiatisé, réflexion sur quelques enjeux" in *Cahiers S.T.S. Ordre juridique, ordre technologique,1996, N° 12, p.* 81 et dans le même *Cahiers* : E. Serverin, "La technique comme analyseur institutionnel, les définitions de la jurisprudence à l'épreuve des banques de données juridiques", p.108.

représentative, ou dont la sélection des données a été biaisée à la source, pourra être remise en question.

B. Ouverture des données d'apprentissage

Enfin, il y a aussi la question du statut juridique des données générées par ces algorithmes. Dès lors que ces données sont générées par une administration publique, il s'agit, *a priori*, de données publiques, qui seront donc soumises au régime d'Open Data. De même, si l'administration s'appuie sur un algorithme opéré par un tiers, il est vraisemblable que les données issues de ces algorithmes seront aussi considérées comme des données publiques, puisqu'elles seront ensuite collectées par les administrations qui en ont commandité la création.

Une question plus complexe se pose dans le cas d'un algorithme fondé sur un système d'apprentissage automatique qui a été développé par un tiers, et dont l'apprentissage ne se fonde pas (uniquement) sur des données publiques, mais s'appuie aussi sur des données privées—potentiellement personnelles et/ou confidentielles. C'est le cas, notamment, du '*credit score*' aux Etats-Unis,[35] ou encore du '*social credit*' en Chine[36] : tous les deux sont calculés à partir d'informations fournies par des compagnies privées comme les bureaux de crédit tels que Experian, TransUnion, and Equifax aux États-Unis, et des entreprises comme Alibaba (pour le Sesame credit) ou Wechat (pour le Tencent credit) en Chine, ainsi que des '*data brokers'*, qui collectent des informations sur les Internautes, et qui les traitent afin de créer des profils d'utilisateurs.

Bien qu'il ne s'agisse pas de données publiques en tant que telles, ces données qui ont été collectées ou générées par des acteurs privés peuvent servir de corpus pour alimenter certains des algorithmes utilisés par les administrations publiques. C'est le cas de la Chine, notamment, qui s'appuie (en partie) sur le Sesame credit de Alibaba pour déterminer le score de ces citoyens.

[35] Jilian McLaughlin (2016), "Prise de décision par les données: l'application des Big Data dans le cadre des services financiers" dans Bourcier, D., De Filippi, P. (eds.) (2016) *Open Data & Big Data : Nouveaux défis pour la vie privée.* Mare & Martin

[36] Anne SY Cheug and Clement YX Chen, (2018) "The rise of the Data State: The case of China's Social Credit System" *Working Paper, 2018.*.

Mais quel régime s'applique alors pour ces données privées? Ces données doivent-elles retomber dans le régime de l'Open Data dès lors qu'elles sont utilisées pour guider l'apprentissage de certains algorithmes, qui seront ensuite utilisés par une administration publique? Et quelles sont les contraintes ou les limitations liées à l'ouverture de ces données?[37]

Une vrai transparence des algorithmes implique un accès aux données fournies *à* ces algorithmes. Mais dans beaucoup de cas, on n'a pas accès à ces données, car celles-ci ne sont pas soumises au régime général de l'Open data. En effet, alors qu'on peut exiger la transparence des algorithmes utilisés par les administrations publiques - aussi bien dans le cas où ces algorithmes sont déployés par les administrations elles-mêmes, que lorsqu'ils sont opérés par des tiers - ils est difficile de revendiquer une ouverture des données sur lesquelles ces algorithmes se fondent, surtout si ces algorithmes ont été entraînés par des tiers. La loi ne prévoit pas la possibilité de demander l'ouverture des données collectées ou produites par des acteurs privés, d'autant plus que ces données (même anonymisées) sont souvent corrélées avec des informations personnelles ou confidentielles, ou relèvent du secret industriel.

Conclusion

L'Open Data exige une ouverture des données et des documents administratifs ainsi que du code source des logiciels, dans le but de garantir l'accès et la réutilisation de toutes les informations collectées ou générées par les administrations publiques. Cette optique de mise à disposition des données publiques se situe au sein d'une politique plus large de "gouvernement ouvert" qui vise à garantir une plus grande transparence des activités de l'État. Cependant, alors que de plus en plus de données sont collectées, catégorisées, analysées, traitées ou même générées par des procédés informatiques, les administrations publiques s'appuient massivement sur des algorithmes pour gérer les risques de son action. Ainsi, l'Open Data ne suffit plus, à lui seul, à garantir les prérogatives d'un gouvernement ouvert. Il faut désormais fournir de la visibilité sur le fonctionnement des algorithmes qui vont traiter ces données.

[37] Ce sont les mêmes questions sur la disponibilité des données d'expérience qui se posent pour vérifier les résultats scientifiques publiés dans les articles soumis à des revues.

Le principe de transparence des algorithmes doit donc être appliqué à toutes les *données opérationnelles* ou *intermédiaires (*et non plus seulement descriptives) qui sous-tendent ces algorithmes.

Bien que le problème n'ait été jusqu'à présent rencontré que dans un nombre de cas limités (*Compas* aux États-Unis et l'*Harm Assessment Risk Tool* au Royaume-Unis), il est vraisemblable que, avec l'arrivée en force du Big data, les administrations publiques vont bientôt être confrontées à ces questions de manière plus substantielle. Il nous paraît alors nécessaire, d'un point de vue prospectif, d'identifier les solutions potentielles qui permettraient de remplir les conditions opérationnelles d'un gouvernement ouvert.

Aux principes énoncés par la loi relative aux conditions que doivent remplir ces données publiques (*fiabilité*, *disponibilité* et *sécurité)*, nous considérons important d'y ajouter le principe d'*exhaustivité* et de *représentativité* de la base de connaissances sur laquelle les algorithmes ont été entraînés. A l'heure du Big data, où la question de l'origine et de l'échantillonnage des données devient fondamentale, de nombreuses discussions pourront être évitées sur la pertinence et la légitimité des décisions administratives automatisées.