HAL
open science

La science des données numériques au service du contrôle fiscal français - Réflexions sur l'Algocratie

Caroline Lequesne Roth

To cite this version:

Caroline Lequesne Roth. La science des données numériques au service du contrôle fiscal français - Réflexions sur l'Algocratie. A. PARIENTE (dir.), Les chiffres en finances publiques, Paris, éd. Mare & Martin, 2019, pp. 177-193., pp.177-193, 2019. hal-03133110

HAL Id: hal-03133110

https://hal.univ-cotedazur.fr/hal-03133110v1

Submitted on 2 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La science des données numériques au service du contrôle fiscal français Réflexions sur l'Algocratie

Caroline Lequesne Roth,
Maître de conférences à l'Université Côte d'Azur, membre du GREDEG (CNRS – UMR 7321) Programme DL4T et du Centre Perelman de philosophie du droit

> *« La volonté d'étendre à la société toute entière ce que l'on imagine être une organisation scientifique du travail a été le lot commun du capitalisme et du communisme de l'ère industrielle.*
>
> *Elle n'a pas disparu de nos jours, mais seulement changé de forme. Son modèle n'est plus celui des lois de la physique classique, mais celui des algorithmes de l'informatique. »*
>
> A. Supiot, La gouvernance par les nombres.

1. L'usage de la science des données au sein des administrations n'est pas nouveau. Cette grille de lecture mathématico-technique de notre réalité, dont la statistique fut historiquement l'une des composantes majeures, constitua un vecteur de transformation décisif de l'État moderne[1]. La prégnance du nombre, « ultime garant de la réalité »[2], est aujourd'hui consubstantielle de nos politiques publiques : « [p]ersonne ne saurait parler sérieusement de l'état de notre société et discuter des politiques sans se référer aux informations quantitatives délivrées par des organismes, institutions, agences spécialement dédiées à leur production »[3] relève le mathématicien philosophe Oliver Rey. Le règne des experts, la multiplication des indicateurs ou l'autorité acquise par l'information quantitative en sont autant de symptômes. La mobilisation de la science des données connaît toutefois, aujourd'hui, des développements inédits. Ceux-ci tiennent tout autant à la nature, numérique, qu'au traitement, algorithmique, des

1. Voy. notamment : I. Bruno & E. Didier, *Benchmarking, l'État sous pression statistique*, Paris, éd. Zones, 2013, 250 p.
2. O. Rey, *Quand le monde s'est fait nombre*, Paris, éd. Stock, 2016, p. 8.
3. *Ibidem*.

memchiffres_finances_publiques-mqt02-ter.indd 177 11/06/2019 16:35:37

données. La diversification de l'information et le renouvellement des méthodes initient ainsi une nouvelle ère de la collaboration technico-administrative, au sein de laquelle certains décèlent les prémisses d'une « algocratie »[4]. Dans le domaine du contrôle fiscal, le recours aux algorithmes dans le traitement des données numériques est en plein essor. Le phénomène est global, et s'est largement répandu au cours de la dernière décennie. Mentionnons à ce titre que l'administration belge a engagé ses premiers travaux dès 2000 pour lutter contre les carrousels de TVA ; le fisc italien s'est doté du « *redditometro* » – l'« indicateur de richesse » – à compter de 2013, pour identifier les inadéquations entre les déclarations de revenus et les habitudes de vie de ses contribuables ; le Royaume-Uni a quant à lui généralisé l'automatisation des contrôles fiscaux à partir de 2010 au travers du dispositif Connect développé par son administration fiscale. Une étude réalisée en 2014 par l'OCDE indiquait que 69 États, parmi les 86 interrogés, avaient recours à des outils de traitement des données numériques dans le domaine fiscal[5]. La France n'échappe pas au phénomène : les administrations financières, et plus particulièrement la Direction Générale des Finances publiques, se sont dotées d'outils de *data mining* pour lutter contre la fraude (I). Le recours à ces outils n'est pas neutre et suscite une transformation profonde de notre droit fiscal (II). La présente contribution dresse un premier bilan de l'usage de ces dispositifs techniques en France, afin de réfléchir aux conditions démocratiques du déploiement d'une administration algorithmique.

I. Le contrôle fiscal à l'ère du *data mining*

2. *Data mining* : essai de définition. La science des données est par nature transdisciplinaire : au croisement des mathématiques et de l'informatique, elle se décline dans une batterie d'outils du *machine learning* au *web scraping*, dont toutes les potentialités ne sont pas encore pleinement exploitées. L'administration française recourt à titre principal au *data mining* qu'elle définit comme « une démarche méthodologique rigoureuse développée en vue de révéler de l'information contenue dans les systèmes d'information, en mettant en exergue d'éventuelles corrélations significatives entre les données observées. »[6] Le *data mining* est en d'autres termes une technologie d'exploitation des données numé-

4. P. Gueydier (dir), *Pouvoir régalien et algorithmes, vers l'algocratie ?*, Optic, janvier 2018, 34 p.
5. Citée par D. Restrepo Amariles & G. Lewkowicz, « De la donnée à la décision : comment réguler par les données et les algorithmes », dans *Les big data à découvert*, Paris, CNRS Editions, 2017.
6. Délégation Nationale à la Lutte contre la Fraude (DNLF), *Le « data mining », une démarche pour améliorer le ciblage des contrôles*, Ministère de l'Économie et des Finances, 14 janvier 2014.

memchiffres_finances_publiques-mqt02-ter.indd 178 11/06/2019 16:35:37

riques visant à « faire parler l'information »[7] en établissant des corrélations entre les données observées.

3. Les administrations françaises à l'heure du *data mining*. Les outils de *data mining* sont aujourd'hui largement mobilisés par les administrations françaises dans la lutte contre la fraude. La Caisse nationale des Allocations familiales (Cnaf) fut de pionnière : ses premiers dispositifs furent introduits en 2006 pour détecter les cas de fraude aux prestations servies par la branche famille ; leur usage a été généralisé à l'ensemble du réseau des caisses en 2011[8]. En 2013, une mission de coordination interministérielle a été confiée à la Direction Nationale à la Lutte contre la Fraude (ci-après DNLF) pour améliorer les dispositifs existants ; elle rassemble l'ensemble des administrations financières et des organismes sociaux qui ont depuis lors multiplié leurs travaux et expérimentations en matière d'exploitation des données. Mentionnons à ce titre les travaux la Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS) sur le champ de la fraude administrative des transporteurs sanitaires[9] ; la mission de Pôle emploi visant à identifier des profils de fraudeurs en matière d'allocation chômage[10] ; les travaux relatifs à l'identification des risques de fraude en matière de travail dissimulé menés par l'Agence centrale des organismes de Sécurité sociale (ACOSS)[11]. En matière douanière, les premiers travaux intéressaient deux missions : la lutte contre la fraude fiscale sur internet et la recherche de marchandises prohibées[12] . Les services de la Direction générale des douanes et droits indirects (DGDDI) ont en outre renforcé leur pôle d'expertise avec la création du service d'analyse de risque et de ciblage (SARC) en juin 2016[13]. Les techniques de *machine learning* (système auto-apprenant) ont également été mobilisées, dans les travaux conduits par la Caisse nationale d'assurance vieillesse (CNAV) entre juin et octobre 2016[14]. La France entend plus largement se positionner en leader sur la scène européenne et internationale, en consolidant les partenariats extérieurs

7. L'expression est courante, on la retrouve notamment dans les communications la DGFiP.
8. A. Chevallier et G. Tauber. « Big data et protection sociale : au-delà de la lutte contre la fraude, des opportunités à saisir pour améliorer l'accès aux droits », *Regards,* vol. 52, no. 2, 2017, pp. 205-215.
9. DNLF, Bilan 2013, *Lutte contre la fraude*, Ministère de l'Économie et des Finances, 2014, p. 16.
10. *Ibidem*.
11. *Ibidem*.
12. DNLF, Bilan 2015, *Lutte contre la fraude*, Ministère de l'Économie et des Finances, 2016, p. 28.
13. Arrêté du 29 février 2016 portant création d'un service à compétence nationale dénommé « service d'analyse de risque et de ciblage ».
14. H. Cyrille, « Le *big data* à l'Agence centrale des organismes de sécurité sociale (ACOSS). Du concept à l'expérimentation », *Revue française des affaires sociales*, pp. 141-149.

memchiffres_finances_publiques-mqt02-ter.indd 179 11/06/2019 16:35:37

sur ces questions avec l'OCDE[15] ou au sein du groupe de travail dédié à la lutte contre le travail dissimulé piloté par la Commission européenne et dont elle assure la vice-présidence[16].

4. Le *data mining* au sein de la DGFiP. La Direction générale des finances publiques fait figure de moteur dans la lutte contre la fraude fiscale. En juillet 2013 fut créée en son sein une mission dédiée au développement du *data mining* nommée « Requêtes et valorisation » et composée de fiscalistes et de *data scientits*. Ses premiers travaux concernaient la lutte contre la fraude au remboursement de crédit de TVA des petites entreprises. Les résultats, encourageants, ont justifié d'élargir l'assiette des contrôles : sur avis favorable de la CNIL[17], un arrêté de 2014 a autorisé l'implémentation d'un dispositif pilote visant à améliorer le ciblage des entreprises fraudeuses[18]. Les modélisations expérimentales concernèrent l'impôt sur les sociétés, la taxe sur les salaires et la détection les entreprises présentant un risque en matière de recouvrement de créances. Pérennisée, l'expérience a été étendue à la recherche des fraudes perpétrées par les particuliers en 2017[19]. Un bilan de l'expérimentation, programmé au printemps 2019, doit statuer sur sa reconduction.

5. Le ciblage en amont des contrôles. En pratique, les dispositifs techniques instaurent un mécanisme de contrôle à double détente : les logiciels participent, en amont, à la détection automatique des fraudeurs ; il appartient dans un second temps aux contrôleurs, au regard des éléments du dossier, de juger de l'opportunité des poursuites.

6. Modélisation algorithmique : le profilage des fraudeurs. Le dispositif technique en amont des contrôles est un système d'analyse prédictive de la fraude, réalisé sur le fondement d'indicateurs statistiques au départ de dossiers déjà instruits ; il permet une modélisation algorithmique des fraudes et des

15. La DNLF travaille en étroite collaboration avec l'OCDE sur la question de l'analyse stratégique des données par les administrations. DNLF, Bilan 2016, *Lutte contre la fraude*, Ministère de l'Économie et des Finances, 2017, p. 28.
16. Au sein de la plateforme européenne de lutte contre le travail non déclaré créé en 2016. La France en assure la vice-co-présidence aux côtés de la Lettonie. Les travaux en *data mining* des administrations sont mobilisés au service de la lutte contre la fraude. *Ibidem*, p. 14.
17. Délibération n° 2014-045.
18. 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 la fraude et valorisation des requêtes ».
19. Avis de la CNIL, Délibération n° 2016-286 ; 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 la fraude et valorisation des requêtes ».

memchiffres_finances_publiques-mqt02-ter.indd 180 11/06/2019 16:35:37

trafics. En 2014, la DNLF a élaboré une méthodologie « cadre » qui identifie les différentes étapes du profilage[20].

L'élaboration des profils de fraudeurs requiert au préalable un paramétrage ; sont nécessairement renseignés : la cible du contrôle, son périmètre et les données à partir desquels il sera perpétré. Le premier élément détermine l'objectif poursuivi. Il se traduit et se décline en termes d'indicateurs, lesquels correspondent aux « variables à expliquer en fonction de facteurs d'influence à tester ». Parmi ceux-ci : la probabilité de procéder à un redressement lors d'un contrôle, le montant du redressement en valeur absolue, le montant du redressement rapporté au montant des cotisations ou de l'impôt dû, le nombre d'allocataires détectés en situation de fraude ou encore le montant des prestations indues stoppées avant versement[21]. La détermination de l'objectif permet en outre l'évaluation a posteriori du modèle établi. Le second paramètre à considérer, le périmètre du ciblage, permet définir le champ d'investigation et de caractériser la fraude. Le périmètre peut être fonction du secteur d'activité, de la zone géographique, de la catégorie d'impôt, du segment de cotisants ou encore du type d'allocataire. Parce qu'ils délimitent la population à examiner, ces paramètres sont critiques en termes d'efficacité. La DGFN observait en effet que les résultats sont plus probants avec des populations considérées comme homogènes[22] ; elle recommandait aussi, en phase expérimentale, de circonscrire le ciblage à des « catégorie(s) spécifique(s) d'individus ». Enfin, la modélisation suppose la constitution d'un entrepôt de données exploitables et fiables. La DGFiP travaille actuellement à la constitution d'un « silo », qui intégrera à terme, de façon décloisonnée, données structurées et non structurées des contribuables. La CNIL impose la recension des fichiers exploités à ces fins ; ils rassemblent, à ce jour, les données propres à l'administration[23], les déclarations transmises par les plateformes numériques collaboratives[24], les informations

20. Délégation Nationale à la Lutte contre la Fraude (DNLF), *Le « data mining », une démarche pour améliorer le ciblage des contrôles*, Ministère de l'Économie et des Finances, 14 janvier 2014.
21. *Ibidem*, p. 3.
22. « [L]'hétérogénéité des comportements de fraude est d'autant plus importante que l'on raisonne sur des populations différentes. Il est dès lors plus délicat de cerner les critères de risque qui soient valables pour l'ensemble de la population sur des périmètres flous ou trop vastes », *ibidem*.
23. Outre les déclarations d'impôts, le silo rassemblerait vingt-deux fichiers rassemblés par Bercy concernant : les revenus transmis par les employeurs, les informations relatives aux comptes bancaires et aux contrats d'assurance-vie, les actes notariés, les données patrimoniales, cadastrales et immobilières, les liens d'associés et de dirigeants d'entreprises, ainsi que le détail des impôts payés et des contentieux en cours. Le fisc peut également solliciter les renseignements auprès d'autres services tels que l'Urssaf ou l'inspection du travail. V. Nouzille, « Comment le fisc vous espionne », *Le Figaro Magazine*, 25 janvier 2019.
24. La loi relative à la lutte contre la fraude fiscale du 23 octobre 2018 impose à ces dernières une obligation de déclaration.

memchiffres_finances_publiques-mqt02-ter.indd 181 11/06/2019 16:35:37

relatives aux transactions de cryptoactifs[25] et aux données bancaires et fiscales étrangères[26]. La presse s'est également fait l'écho de la collecte de données issues des réseaux sociaux ; le ministre de l'action et des comptes publics lui-même a fait état de la volonté de recourir aux méthodes de *webscrapping* pour exploiter celles-ci[27]. Pour l'heure, ces perspectives demeurent hypothétiques, l'avis de la CNIL, non sollicité, étant entre autres requis pour engager de telles démarches.

Le paramétrage réalisé, l'exploration des données en vue du ciblage des fraudeurs s'effectue en trois étapes. Premièrement, l'analyse descriptive des caractéristiques des populations concernées permet d'identifier les critères de risque de fraude. Ils constituent le fondement des modèles économétriques réalisés dans un second temps : ces derniers, contrôlés et réajustés, correspondent au profil type des fraudeurs. Ils sont dans un troisième temps appliqués sur une plus large population pour repérer les profils à risque auxquels est assigné un score renseignant la probabilité de fraude.

7. Modélisation algorithmique : l'identification des réseaux. Au profilage des fraudeurs s'ajoute une cartographie des réseaux de fraude. La DGFIP a sollicité les services de deux sociétés spécialisées dans ladite « théorie des graphes »[28] afin d'établir les relations occultes entre individus, sociétés et conseils qui caractérisent ces réseaux. Elle s'est ainsi dotée d'un double outil de lecture de la fraude, alliant à une approche « micro » et comportemental, un outil de compréhension « macro » et relationnel/transactionnel. L'opération a vocation à cibler l'individu dans son environnement.

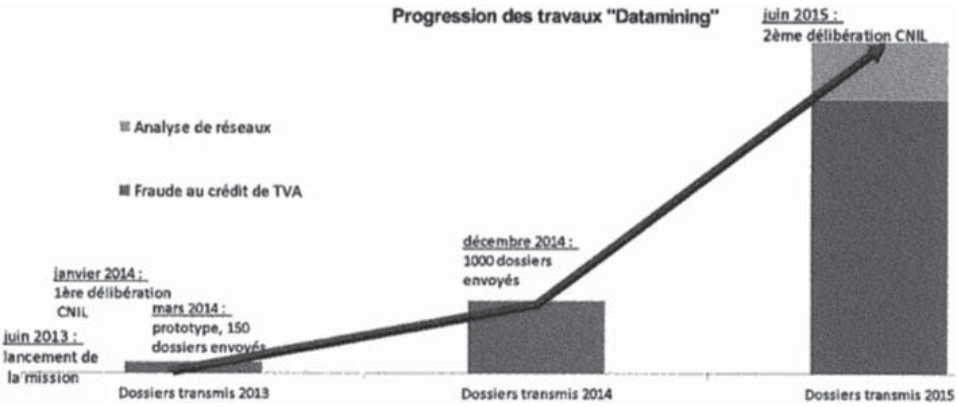
8. Bilan des dispositifs français. L'expérience française est aujourd'hui embryonnaire. Les dispositifs doivent encore être déployés à l'échelon régional[29] et différents rapports font état du manque d'agents compétents. L'obstacle le plus décisif tient à la constitution de l'entrepôt des données. Les administrations font face à une double difficulté. D'une part, l'accès à des données fiables demeure encore un défi. L'étude d'impact de la loi relative à la lutte contre la fraude fiscale indique qu'en l'état actuel des développements, « les États ne sont pas en mesure

25. La loi de finances 2019 impose de déclarer les comptes en cryptoactifs détenus auprès de plateformes étrangères.
26. Conformément aux conventions internationales et aux dispositions européennes concernant les échanges d'informations fiscales conclues depuis 2016.
27. D. Leloup, « Des données des réseaux sociaux utilisées par le fisc : pourquoi ce n'est pas si simple », *Le Monde*, 12 novembre 2018.
28. La DGFIP a en effet eu recours aux services de la société Neo4J et Linkurious qui permettent de modéliser les réseaux, c'est à lire identifier de manière intelligible les réseaux complexes. E. Paquette, « Impôts : pourquoi le logiciel antifraude de Bercy interroge », *L'Expansion – L'express*, 17 septembre 2018.
29. Document de politique transversale projet de loi de finances 2019, Lutte contre l'évasion et la fraude fiscales, ministère de l'action et des comptes publics, p. 37.

memchiffres_finances_publiques-mqt02-ter.indd 182 11/06/2019 16:35:37

(…) d'obtenir des données fiables concernant l'identification des contribuables et des revenus imposables »[30], eu égard, notamment, au défaut de coopération des plateformes. Dans le même sens, l'assiette des données exploitées par la DGFiP est plus restreinte que la CNIL ne l'autorise en raison de la qualité insuffisante des données. D'autre part, une large partie des informations d'ores et déjà en possession de l'administration sont à ce jour inexploitables. Les données ne sont pas liquides par nature : elles requièrent un traitement de mise en forme dit « fusion de données » qui assure leur caractère homogène et opératoire. Ce travail, fastidieux, n'est pas abouti et de nombreux renseignements, à l'instar des données en provenance de l'étranger, ne seraient pas traités à ce jour[31].

Pour autant, les premiers résultats obtenus sont qualifiés d'« encourageants » par la DNLF. Ils traduisent un rendement croissant des contrôles. Les plus probants concernent la lutte contre les fraudes en matière de TVA. Comme le montre le schéma ci-dessous, le taux de dossiers transmis ayant fait l'objet d'une programmation est en progrès constant.



Source : Rapport remis par le Gouvernement au Parlement en application de l'article 24 de la loi n° 2014-1655 du 29 décembre 2014 de finances rectificatives pour 2014, p. 23.

Le bilan des premières expérimentations avançait en outre que 35 % des dossiers repérés par le logiciel ne l'avaient pas été par les contrôleurs. Les mêmes constats ont été dressés concernant les fraudes des entreprises[32]. Selon les chiffres

30. Étude d'impact, Projet de loi relatif à la lutte contre la fraude, 27 mars 2018, p. 26.

31. La presse rapporte ainsi que les listings de comptes bancaires étrangers appartenant à des Français recueillis depuis 2017 par Bercy, qui représentent « plusieurs millions de données » n'ont pas encore été exploités. V. Nouzille, « Comment le fisc vous espionne », précédemment cité.

32. « L'identification de ces affaires n'aurait pas été possible avec les méthodes traditionnelles de programmation » Document de politique transversale projet de loi de finances 2019, précédemment cité, p. 36 ; ou encore : « Les services ont été destinataires, fin

memchiffres_finances_publiques-mqt02-ter.indd 183 11/06/2019 16:35:37

avancés par la DGFiP, en 2017, la mission requête et valorisation a transmis aux services de contrôle 27 listes portant sur des problématiques professionnelles ; au 31 mars 2018, ces productions avaient conduit à l'examen de 28 857 dossiers[33] ; de janvier à août 2018, ils ont été à l'origine de plus de 18 000 contrôles du bureau ayant donné lieu à plus de 55 M€ de rappels de droits et pénalités (contre 44 M€ en 2017) et à la programmation d'environ 4 300 contrôles fiscaux externes (contre 3 000 en 2017)[34].

Ces résultats sont corroborés par ceux d'autres administrations ayant également recours à du traitement automatisé de données. À titre d'exemple, la CAF a enregistré une augmentation de 46 % du rendement de ses contrôles[35] tandis que 99,6 % des offres d'emploi frauduleuses ont pu être identifiées par l'algorithme de détection de Pôle Emploi[36].

9. Les perspectives françaises : renforcement et optimisation des contrôles automatiques. Les annonces gouvernementales et les budgets engagés au service du développement de ces dispositifs laissent à penser que l'automatisation du ciblage en matière de fraude fiscale n'en est qu'à ses prémisses. Le ministre de l'action et des comptes publics affiche des objectifs ambitieux : « [e]n 2018, 15 % des contrôles ont été programmés grâce à cet outil de *data mining*. Nous espérons que cette part montera à 50 % d'ici à un an »[37]. À cet effet, le budget de la mission requête et valorisation a été augmenté de 20 M€ au titre de l'année 2019. La DGFiP projette de transformer celle-ci en « service national d'analyse de données opérationnel », dont les travaux à l'origine des opérations de contrôle doivent « progresser significativement » dans les années à venir[38]. Dans le même sens, il est intéressant d'observer que le gouvernement envisage de remédier à la longueur des procédures de contrôle, engendré par l'instruction « manuelle » des dossiers, en recourant à ces méthodes. Le traitement algorithmique des pièces est présenté comme une solution d'optimisation à privilégier[39].

2016, de plusieurs milliers de dossiers présentant des anomalies ou des incohérences qui n'auraient pu être détectées qu'au terme de consultations manuelles et répétitives des applications informatiques », DNLF, Bilan 2016, *Lutte contre la fraude*, Ministère de l'Économie et des Finances, 2017, p. 28.

33. *Ibidem*.

34. Document de politique transversale projet de loi de finances 2019, précédemment cité, p. 36.

35. A. Chevallier et G. Tauber. « *Big data* et protection sociale : au-delà de la lutte contre la fraude, des opportunités à saisir pour améliorer l'accès aux droits », précédemment cité.

36. DNLF, Bilan 2017, *Lutte contre la fraude*, précédemment cité.

37. V. Nouzille, « Comment le fisc vous espionne », précédemment cité.

38. DNLF, Bilan 2017, *Lutte contre la fraude*, précédemment cité, p. 29.

39. Voy. en ce sens Rapport remis par le Gouvernement au Parlement en application de l'article 24 de la loi n° 2014-1655 du 29 décembre 2014 de finances rectificatives pour 2014.

memchiffres_finances_publiques-mqt02-ter.indd 184 11/06/2019 16:35:37

10. Les perspectives à l'échelon global : l'exemple de Connect. Les expériences étrangères, et plus particulièrement l'expérience britannique, offrent une projection de ce que pourrait, à terme, constituer le modèle français. Notons que l'automatisation partielle des contrôles fiscaux, lancée en 2010 pour remédier aux pertes engendrées par la fraude, a requis un investissement conséquent. Outre une évolution de la législation fiscale pour permettre l'introduction de cet outil, son coût financier s'élève à 90 M£[40]. Au sein du HMRC – *Her Majesty Revenue & Customs*, l'administration fiscale britannique – la cellule dédiée au dispositif regroupe 250 analystes et plus de 1 000 employés[41]. L'état d'avancement des travaux et l'expertise acquise se mesurent dans l'élaboration de l'entrepôt de données. Le maillage informationnel du fisc britannique, issu de plus de 30 sources différentes, est extrêmement dense[42]. Il permet, en sus un ciblage exhaustif, une revue instantanée des contribuables suspects. L'administration a notamment su exploiter à bon compte les données étrangères ; depuis 2016, Connect se construit en réseau[43]. Le dispositif serait aujourd'hui à l'origine de 98 % des contrôles perpétrés. Selon le HMRC, Connect aurait rapporté 3 milliards £ en huit ans et réduit le manque à gagner fiscal de 8,3 % (2004-2005) à 6,5 % (2014-2015)[44]. La Belgique se prévaut également de résultats concluants : les outils de *data mining* adoptés pour lutter contre la fraude à la TVA aurait permis, entre 2001 et 2010, une baisse 80 et 90 % de celle-ci, pour un montant estimé à près d'un milliard d'euros[45].

II. La transformation algorithmique du droit fiscal

11. L'hypothèse de la technologie ancillaire. Les dispositifs algorithmiques transforment-ils notre droit ? Les positions adoptées au sein de la doctrine française tendent à écarter cette hypothèse, certains auteurs décelant dans l'automatisation

40. Et un milliard supplémentaire devrait être investi au titre de l'année 2019. HMRC, Connect Briefing Public, consulté le15 février 2019, à l'adresse suivante : https://www.tax.org.uk/policy-technical/technical-news/compliance-reform-forum-slide-packs-december-meeting.
41. European Platform Undeclared Work, Good practice fiche – UK : Data mining tools and methods to tackle the hidden economy in the UK, 2017.
42. Regroupant au total plus d'un milliard de données structurées. R. Dyson, « What does the taxman know about you, your finances and your lifestyle ? » *Telegraph*, 25 Jun 2015.
43. Depuis 2017, Connect peut accéder aux données de plus de 60 pays.
44. European Platform Undeclared Work, Good practice fiche – UK, précédemment cité.
45. IFRAP, « *Big data* : en matière de contrôle fiscal, la France va devoir mettre les bouchées doubles », 15 mai 2014, consulté le 15 février 2018 à l'adresse suivante : http://www.ifrap.org/etat-et-collectivites/big-data-en-matiere-de-controle-fiscal-la-france-va-devoir-mettre-les-bouchees.

memchiffres_finances_publiques-mqt02-ter.indd 185 11/06/2019 16:35:37

des procédures « la garantie d'une application uniforme de la loi », prémunissant « les administrés contre les erreurs des agents »[46] voire « l'arbitraire administratif »[47]. Les algorithmes et modélisations mathématiques, réduits aux confins d'une technologie ancillaire, optimiseraient le travail de l'administration dans le respect des pouvoirs établis. Cette compréhension du phénomène est confortée par la décision du Conseil Constitutionnel du 12 juin 2018 concernant la loi relative à la protection des données personnelles. La Haute juridiction française était invitée à se prononcer sur l'extension des hypothèses dans lesquelles l'administration peut recourir à une décision administrative sur le seul fondement d'un traitement automatisé de données. Selon les sénateurs requérants, l'appréciation des situations individuelles par le truchement algorithmique – et à plus forte raison, par les algorithmes « autoapprenants » – méconnaissait les garanties constitutionnelles inhérentes à l'exercice du pouvoir réglementaire. Les sages écartèrent toutefois cette lecture, arguant de la neutralité des dispositifs techniques sur l'ordre juridique national : ceux-ci « n'ont ni pour objet ni pour effet d'autoriser l'administration à adopter des décisions sans base légale, ni à appliquer d'autres règles que celles du droit en vigueur ». Le réceptacle technique n'altère ainsi nullement contenu de la règle de droit, dont la détermination – et l'application même – lui sont étrangère[48].

12. Au-delà de la neutralité technologique. Une analyse attentive des effets produits par les dispositifs algorithmiques appelle toutefois à la nuance. Trois observations interrogent les fondements mêmes de cette approche : la place accordée à l'humain dans le processus décisionnel, la concurrence des modèles – ou « pattern » – à la règle de droit et la diffusion des standards processuels « embarqués » dans l'objet technique.

13. Performativité et légitimation algorithmiques. Les discours et scénarios dystopiques décrivant le remplacement de l'humain par le robot ont abondamment nourri les sciences sociales. Si l'idée de la « substitution logicielle »[49] n'est pas neuve, sa réalité, aujourd'hui tangible, laisse entrevoir les contours d'un « Léviathan algorithmique »[50]. Les systèmes de lutte contre

46. E. Untermaier-Kerléo, « Les nouveaux visages de la décision administrative : d'une administration assistée à une administration automatisée », *La Semaine Juridique Administrations et Collectivités territoriales,* n° 50, 17 décembre 2018.
47. *Ibidem*.
48. Notons que ce même postulat est au fondement du système Compas utilisé par un nombre croissant de juridictions pénales pour statuer sur les demandes de remise en liberté. Voy en ce sens Cour suprême du Wisconsin, State of Wisconsin v. Eric L. Loomis, 13 juillet 2016, case n° 2015AP157-CR.
49. E. Sadin, *L'intelligence artificielle ou l'enjeu du siècle : Anatomie d'un antihumanisme radical*, Paris, éd. L'Echappée, octobre 2018, p144.
50. *Ibidem,* p.133.

memchiffres_finances_publiques-mqt02-ter.indd 186 11/06/2019 16:35:37

la fraude, et à plus forte raison contre la fraude fiscale, en constitue à notre sens une des plus vives illustrations. Elle se caractérise en premier lieu dans la réduction de l'opportunité des contrôles qui s'opère au bénéfice du dispositif. Bien que l'administration défende la prévalence des agents dans l'évaluation de l'opportunité des poursuites, les scores obtenus au terme du ciblage ne sont pas sans effet. Des études conduites à Harvard pour évaluer l'impact d'outils d'analyse prédictive dans l'élaboration des décisions judiciaires confirme l'analyse[51]. Elle montre que la mise en cause des résultats des logiciels de prédiction par les juges est rare, au regard de la présomption de justice dont bénéficient les dispositifs techniques[52]. De même, « l'acceptabilité » des décisions judiciaires par le justiciable serait plus forte lorsque fondées sur un protocole automatisé[53]. Il en irait ainsi au civil, en matière indemnitaire, ou au pénal comme l'illustre la réception du *scoring* de récidive utilisé par nombre de juridictions américaines[54]. Le phénomène s'explique à l'aune des schèmes de légitimation sur lesquelles reposent nos structures de pouvoir : la rationalité scientifique, dans l'illusion du paroxysme algorithmique, apparaît incontestable. En d'autres termes, la performativité des résultats produits apparaît d'autant plus délicate à contester qu'elle matérialise les soubassements théoriques du pouvoir. La rationalité algorithmique se substitue ainsi à l'état de droit comme vecteur de confiance. La délégation est manifeste dans le cas du HMRC, où Connect est à l'initiative de la quasi-totalité des contrôles diligentés ; elle est subreptice dans le cas français, mais n'en demeure pas moins l'horizon désiré, comme en attestent les ambitions affichées.

13a. Technogouvernementalité. La conjoncture s'avère de surcroît propice à la délégation technologique organisée au nom de la gestion des deniers publics. L'orthodoxie budgétaire, dominante dans les démocraties occidentales, appelle à la maîtrise de la dépense publique et se traduit, notamment, dans la volonté politique de réduction du nombre de fonctionnaires. Si l'objectif n'est pas expressément affiché en France, force est de constater qu'il a déjà été mis en œuvre à l'étranger. L'État du Michigan en constitue un bon exemple : au sein de l'agence pour l'emploi, l'adoption d'un dispositif de décisions administratives automatisées a

51. J. M. Logg, J. A. Minson, Don A. Moore, « Do People Trust Algorithms More Than Companies Realize? », Harvard Business Review, October 26, 2018, consulté le 15 février 2019, à l'adresse suivante : https://hbr.org/2018/10/do-people-trust-algorithms-more-than-companies-realize. (*Harvard Business Review*).
52. Antoine Garapon évoque à cet égard la « pression » exercée par la justice prédictive.
53. « Le justiciable, par goût du confort, de la rapidité et de la sécurité, est plus enclin à accepter la décision d'un juge robot que celle d'un juge humain » Propos d'Antoine Garapon à l'occaison du colloque « Justice prédictive : évolution, révolution ? » organisé le 23 mai 2017 par la Cour d'appel de Paris.
54. A. Garapon & J. Lassègue, *Justice digitale. Révolution graphique et rupture anthropologique*, Paris, éd. PUF, avril 2018, p.252.

memchiffres_finances_publiques-mqt02-ter.indd 187 11/06/2019 16:35:37

conduit au licenciement d'environ 400 employés[55]. La culture de la performance produit en outre une pression de rendement sur les fonctionnaires en exercice. L'automatisation partielle des tâches et contrôles de l'administration constitue à cet égard une solution accommodante : l'assistance algorithmique au service de l'efficacité administrative. Pour autant, son développement renforce in fine la contrainte exercée sur les agents, en ce qu'elle porte en germe la condition même de leur activité : comment assurer le rendement attendu de cet investissement d'équipement sans y recourir ? Comment justifier une productivité « humaine » à l'heure du traitement automatisé ? Les dispositifs techniques, et plus spécifiquement algorithmiques, de l'administration présentent à cet égard toutes les caractéristiques des instruments de « techno gouvernementalité » : loin de simples vecteurs informationnels d'assistance, ils conditionnent les jugements et orientent les conduites.

14. Le droit en forme de modèle. En second lieu, notons que le profilage des fraudeurs auxquels conduisent les systèmes apprenants de *data mining* transforme substantiellement la nature de notre droit : la règle se mue en modèle, le modèle étant lui-même compris comme une corrélation de faits. Ce processus n'est pas propre aux algorithmes de lutte contre la fraude fiscale[56], mais se révèle de manière particulièrement prégnante au travers d'eux. Comme le décrivent les schémas de fonctionnement reproduits ci-dessus, le fraudeur n'est pas tant ciblé à partir de l'infraction commise qu'au départ de signaux, comportementaux et/ou relationnels, identifiés itérativement par le système. Le profil type du fraudeur est un modèle construit au départ des faits et des correspondances factuelles ; le scoring, qui indique la probabilité de fraude et conditionne en partie les poursuites, évalue l'adéquation à ce modèle. Le droit se transforme donc en corrélation de faits et les faits, corrélés, deviennent, indépendamment de leur légitimité, normatifs[57].

14a. De la cause à la corrélation. S'opère ainsi, plus fondamentalement, la migration d'un système juridique causal vers un système de corrélations pratiques[58]. La présomption de fraude procède dans chacun d'eux, de logiques distinctes. Au terme d'un raisonnement juridique causal, l'agent identifie la fraude au regard des éléments constitutifs de celle-ci, en appréciant le contexte

55. Lesquels étaient affectés principalement à la vérification de cas de fraude potentielle. F. Brochu-Lemay, « MiDAS : L'intelligence artificielle devenue juge, jury et bourreau », Blogs, Laboratoire de Cyber Justice, 24 juillet 2018, consulté le 15 février à l'adresse suivante : https://www.cyberjustice.ca/en/actualites/2018/07/24/midas-lintelligence-artificielle-devenue-juge-jury-et-bourreau/

56. Comme l'ont établi les travaux de Gregory Lewkowicz. Voy. par exemple, Leçon inaugurale de la Chaire A. Koyré, *L'émergence de la smart law : le droit dans la société algorithmique*, Nice, le 15 octobre 2018. https://www.droiteconomique.org/chaire-koyre-2018-2019/

57. A. Garapon & J. Lassègue, *Justice digitale, op.cit.* p. 224.

58. *Ibidem*, p. 223. Sur le sujet, voy. Également : D. Cardon, *A quoi rêvent les algorithmes*, Paris, éd. Seuil, octobre 2015, pp.39-44.

memchiffres_finances_publiques-mqt02-ter.indd 188 11/06/2019 16:35:37

dans lequel elle a pu être perpétrée ; il rassemble un faisceau d'indices concordants en procédant à un arbitrage qualitatif des informations dont il dispose. Dans le cas d'une évaluation algorithmique, le logiciel identifie des concordances de faits, discriminantes au regard des dossiers déjà instruits ; il délivre un score de probabilité d'après une masse de données, appréciant la situation de chacun au travers de paramètres quantitatifs. La normativité juridique « sociale et herméneutique »[59] est ainsi décontextualisée au bénéfice d'une vérité scientifique.

Ajoutons que la régulation perpétrée devient *ex ante*, la fraude étant présumée avant même sa réalisation. Ce phénomène ne va pas sans soulever d'importantes interrogations concernant l'état de droit.

15. Standards globaux embarqués. Enfin, l'automatisation des contrôles conduit à l'émergence d'un nouveau type de standards globaux. « Embarqués » dans les dispositifs algorithmiques, ils se diffusent par le truchement des cocontractants de l'administration qui les élaborent. Les travaux du Centre Perelman ont contribué à identifier et analyser le processus à l'œuvre en d'autres chantiers. Les études relatives aux *law firms* décrivent ainsi le rôle acquis par les avocats dans la production des normes globales : des constitutions en kit aux contrats d'emprunt d'État, leurs « produits » juridiques innervent nos systèmes juridiques nationaux. Cela tient à l'extrême concentration du marché aux souverains et l'effet d'échelle qu'elle emporte : dès lors qu'une nouvelle clause intègre un contrat standard adopté par une cinquantaine d'États, cette clause s'impose comme le standard du marché dans lequel elle s'inscrit[60]. Cet effet de levier est également observable concernant les marchés de service technologique. Si l'administration française est attachée au recrutement et à la formation de services *in house* hautement qualifiés, elle n'en échappe pas moins à la pratique. Pour mettre au point ses dispositifs, la DGFiP a sollicité les services de différents opérateurs économiques, dont la société SAS Analytics, avec laquelle elle travaille par ailleurs sur des outils prédictifs d'aide à la décision pour les finances publiques[61]. SAS Analytics est devenue une référence dans le domaine du *data mining* administratif. En matière de détection de fraude, ses services ont notamment été mobilisés par de nombreuses administrations américaines[62] ou le HMRC britannique[63]. En mai 2012, elle

59. *Ibidem*, p. 225.
60. Nous renvoyons notre aimable lecteur à nos travaux : *Le régime contractuel des États débiteurs européens,* Paris, LGDJ, décembre 2018, pp. 25-33.
61. Sur le site de l'entreprise, par La Direction Générale des Finances Publiques, « Des indicateurs d'aide à la décision pour les finances publiques », consulté le 15 février 2019 à l'adresse suivante : https://www.sas.com/fr_ca/customers/temoignages-clients_msm_moved/dgfip.html.
62. Parmi lesquelles le County de Los Angeles https://www.sas.com/en_us/customers/la-county-dpss.html ou l'administration du Michigan, voy. *infra*.
63. Notamment pour le stockage des données, P. Rigney, « The all seeing eye : an HMRC success story ? », *FTA HMRC Administration*, November/December 2016.

memchiffres_finances_publiques-mqt02-ter.indd 189 11/06/2019 16:35:37

scellait un accord avec une autre société de référence, Cap Gemini, à laquelle les douanes françaises ont eu recours pour leur propre dispositif de *data mining* au sein du SARC. L'état de service des deux entités couvrirait l'Europe, le Brésil, l'Inde et l'Asie Pacifique[64]. Cap Gemini a en outre conseillé le secteur public australien, canadien et la Commission européenne pour la construction de son portail de données. De la même manière que certains arguent de la constitution d'une « Loi Google » au travers du droit (émanant) des plateformes[65], se dessine en matière de lutte contre la fraude une « modélisation SAS – Cap Gemini ».

16. Subjectivité technique : de la discrimination algorithmique. Cette standardisation ascendante est d'autant plus inquiétante que les modèles ne sont jamais neutres. Ils constituent, selon la mathématicienne Cathy O'Neil un « ensemble d'opinions inséré dans un système mathématique » qui, « malgré leur réputation d'impartialité, sont le reflet d'une idéologie »[66]. Leur paramétrage relève nécessairement de choix – stratégique, économique, social et in fine politique – qui engendre des biais.

En France, l'examen du code source de la plateforme APB, obtenu par les associations étudiantes à l'issue d'une bataille contentieuse, a contribué à mettre le phénomène en évidence. Plus que biaisé, le dispositif apparaît illégal : il a été établi qu'il opérait des discriminations, notamment sur le fondement de la nationalité[67]. Rappelons que l'article 14 de la Convention européenne des droits de l'homme (CEDH) interdit les discriminations sur le fondement de l'origine nationale[68]. La disposition s'oppose également à la discrimination en raison de la fortune, qui a suscité par ailleurs des critiques envers les outils de *data mining* déployé par la CAF. Un rapport du Défenseur des droits de 2017 dénonce en effet le dispositif de lutte contre la fraude aux prestations sociales, au sein duquel les critères socioprofessionnels apparaissent déterminants. Le rapport met en exergue

64. SAS And Capgemini Tackle Fraud And Improper Payments In Public Sector, Collaboration will help reduce revenue losses for tax and welfare agencies, Amsterdam, May 8, 2012, consulté le 15 février 2019 à l'adresse suivante : https://www.capgemini.com/wp-content/uploads/2017/07/pdf_SAS_and_Capgemini_Tackle_Fraud_and_Improper_Payments_in_Public_Sector.pdf.

65. Voy. en ce sens J. Charpenet, « La corégulation par les plateformes – Le cas des Governemental Request », *R.I.D.E.*, 2019/3, à paraître.

66. C. O'Neil, *Algorithmes : la bombe à retardement*, Paris, éd. Les Arènes, novembre 2018, p. 40.

67. G. Ouattara, « Parcoursup : comprendre les algorithmes au service de l'affectation des étudiants dans le supérieur », dans *Third, n° 1 Qui gouverne les algorithmes ?,* novembre 2018, pp.32-35.

68. La question se pose à nouveau dans le cas de la plateforme Parcoursup qui remplace APB. L'association étudiante UNEF a introduit un recours administratif pour obtenir le code source du logiciel afin d'en tester les biais. Par une décision en date du 4 février 2019, le Tribunal de Terre-Neuve a accédé à cette demande et assigné l'université des Antilles à communiquer le code source.

memchiffres_finances_publiques-mqt02-ter.indd 190 11/06/2019 16:35:37

un ciblage plus fréquent des bénéficiaires qui perçoivent les prestations sociales les plus précaires[69]. Ces résultats s'expliquent au regard des « facteurs de risque » susceptibles de générer un signalement, au titre desquels figurent l'absence ou l'irrégularité d'emploi, et l'absence ou la faiblesse de ressources.

La modélisation conduit inévitablement à la stigmatisation, qui se voit *a fortiori* renforcée par les boucles de rétroaction : les résultats obtenus sur une population ciblée, qui permettent le réajustement et l'optimisation des contrôles ultérieurs, aggravent inévitablement le ciblage de celles-ci. Le passé se projette sur l'avenir en reproduisant d'un déterminisme factice, manifestement contraire à nos droits fondamentaux.

La DNLF a elle-même admis l'existence de « biais de ciblage », qu'elle juge « inhérents aux contrôles »[70]. Le défenseur des droits encourage ainsi le retour au contrôle aléatoire « moins efficace », mais plus respectueux de l'égalité de traitement des bénéficiaires de prestations ; il invite aussi la cellule de coordination à « engager une réflexion sur les alternatives à l'exploitation automatisée des données (*data mining*) afin de mieux garantir l'égalité de traitement des usagers »[71].

En forme de conclusion : l'algorithme au péril de la démocratie, quelques enseignements pour demain

17. Le recours à la science des données et l'adoption d'outils de *data mining* au service de la lutte contre la fraude fiscale éprouvent nos systèmes juridiques et mettent à l'épreuve nos démocraties. À l'heure où les modèles français demeurent encore embryonnaires, trois chantiers majeurs nous semblent devoir être investis pour penser le cadre dialectique dans lequel s'inscriront les liens entre administration et outils algorithmiques.

18. Faillibilité et sécurité des systèmes. En dépit de la confiance placée dans les dispositifs techniques, ces derniers sont susceptibles de présenter de nombreuses failles techniques et/ou sécuritaires. Les expériences étrangères ont montré la nécessité d'évaluer pleinement celles-ci au regard des conséquences hautement

69. En 2014, parmi la population contrôlée, la proportion d'allocataires percevant le revenu de solidarité active (RSA) a été largement supérieure à celle de l'ensemble des allocataires de la branche famille (40,6 points de pourcentage en plus), Défenseur des droits, Lutte contre la fraude aux prestations sociales : à quel prix pour les droits des usagers ? Rapport de septembre 2017, p. 21 consulté le 15 février 2019 à l'adresse suivante :
https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=16746.
70. *Ibidem*.
71. *Ibidem*.

memchiffres_finances_publiques-mqt02-ter.indd 191 11/06/2019 16:35:37

dommageables qu'elles ont engendrées. L'affaire Midas, du nom de l'algorithme mis en place par l'agence pour l'emploi du Michigan (*Unemployment Insurance Agency*, ci-après UIA) afin de détecter les fraudes, en est un cas topique[72]. La spécificité du dispositif tenait à la suppression de supervision humaine, le système d'allocation et de sanction ayant été intégralement automatisé. Les dysfonctionnements furent nombreux : si le dispositif a rapporté des sommes considérables, ce fut au prix de nombreuses erreurs, 93 % des cas de fraudes identifiées et sanctionnées s'avérant erronés[73]. L'automatisation des sanctions privait en outre les administrés de voies de recours effectives, plaçant des publics particulièrement vulnérables en situation de grandes difficultés. Enfin, des failles dans la sécurité des données ont été révélées. La mise à jour du système en 2016 permit à l'ensemble du personnel de l'agence d'accéder librement aux données personnelles des administrés alors même qu'il n'y était pas habilité[74]. Deux catégories de *class action* ont été introduites par les administrés à l'encontre de l'UIA d'une part, des sociétés qui ont élaboré le système d'autre part. Déboutés en appel dans le cadre de la première affaire, les requérants ont introduit un recours, aujourd'hui pendant, devant la Cour Suprême du Michigan. Parmi les cas qui composent la seconde catégorie d'actions, notons que la société SAS Analytics, que l'administration française a également sollicitée, a été condamnée en janvier 2019 par la Cour d'Appel du 6e circuit pour avoir indûment mis un terme aux allocations chômage des plaignants[75]. Une affaire similaire est à déplorer en Australie[76]. Les enseignements de ces affaires sont riches quant à l'expertise technique requise par ces dispositifs. L'efficacité de ces systèmes est à la mesure de leur cécité. Aussi, ils doivent être hautement éprouvés en amont de leur adoption, et régulièrement évalués une fois mis en œuvre.

19. Coûts financiers et humains. La question du coût des dispositifs doit par ailleurs être envisagée sous tous ses angles : financier, humain, écologique. Si la question budgétaire justifie et légitime la délégation algorithmique, sa véracité demande encore à être établie. L'équation est complexe à résoudre et dépasse le cadre de cette contribution. Gardons seulement à l'esprit que les dispositifs – et leurs mises à jour inéluctables – ont un coût financier non négligeable, et que les réductions de personnels se répercutent inévitablement sur les dépenses sociales de l'État. Le coût de la délégation s'évalue également en termes de dépendance

72. F. Brochu-Lemay, « MiDAS : L'intelligence artificielle devenue juge, jury et bourreau », précédemment cité.
73. Comme le rapporte la décision Cahoo citée *infra*.
74. Comme l'UIA le révèle sur son site, consulté le 15 février 2019 à l'adresse suivante : https://www.theregister.co.uk/2017/03/08/ fraud_detection_system_with_93_failure_rate_gets_it_companies_sued/
75. Cahoo v. SAS Analytics Inc., No. 18-1296 (6th Cir. 2019).
76. C. Knaus, « Centrelink scandal : tens of thousands of welfare debts wiped or reduced », *The Guardian*, 13 September 2017.

memchiffres_finances_publiques-mqt02-ter.indd 192 11/06/2019 16:35:37

– voire de capture – de la puissance publique par les opérateurs économiques qui élaborent les dispositifs, les implémentent et les contrôlent. La question de la délégation s'inscrit plus fondamentalement dans de débat relatif à la préservation de l'intérêt général qui devient délicat à assurer dans l'ignorance technologique. Cette question est d'autant plus impérieuse que les dispositifs implémentés bénéficieront en toute vraisemblance d'un « effet cliquet » au regard des investissements suscités.

20. Précarité du système juridique. Enfin, et comme nous l'avons vu, la mobilisation de la science des données n'est pas indolore pour nos systèmes juridiques. Si les outils algorithmiques intègrent nos procédures administratives, ils doivent opérer dans la conscience des acteurs et le respect de l'état de droit. Nous plaidons à cet effet pour l'introduction et la systématisation des audits, et ce à un double niveau : celui des entrepôts de données d'une part, de la modélisation algorithmique d'autre part. Cathy O'Neil préconise, pour leur mise en œuvre, une implication des parties concernées dans l'élaboration des modèles, la réalisation d'études d'impacts approfondies et l'adoption de procédures de contrôle régulier[77]. Paradoxalement, l'intervention humaine n'a jamais été aussi impérieuse : à l'heure de l'algocratie, elle est la condition de la préservation de nos démocraties.

77. C. O'Neil, *Algorithmes : la bombe à retardement*, *op. cit.*, p. 324.

memchiffres_finances_publiques-mqt02-ter.indd 193 11/06/2019 16:35:37