Drague fiscale sur les réseaux sociaux : de nouveaux algorithmes d'apprentissage pour traquer la fraude

D. Desbois

▶ **To cite this version:**

D. Desbois. Drague fiscale sur les réseaux sociaux : de nouveaux algorithmes d'apprentissage pour traquer la fraude. 2019. hal-02406386

HAL Id: hal-02406386
https://hal.science/hal-02406386

Submitted on 12 Dec 2019

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.

Drague fiscale sur les réseaux sociaux : de nouveaux algorithmes d'apprentissage pour traquer la fraude

En France, les estimations de fraude fiscale varieraient entre 2 et 80 milliards d'euros (Md€) selon le rapport parlementaire de Bénédicte Peyrol (LREM)[1] ; ce qui expliquerait l'injonction adressée par le Président Emmanuel Macron à la Cour des comptes, le 25 avril dernier, pour faire la lumière sur cette question controversée dans un contexte de tensions sur les finances publiques et de baisse du consentement à l'impôt. Dans un courrier adressé le 9 mai à Didier Migaud, président de cette institution, le Premier ministre Édouard Philippe indique que *« le moment est venu de dresser un bilan de l'ampleur de la fraude fiscale dans le pays et d'évaluer l'action des services de l'État et les outils qui sont mis en place »*.

Une interview récente[2] de Gérald Darmanin, ministre de l'Action et des Comptes publics, vient de révéler le projet du gouvernement français d'utiliser des algorithmes d'apprentissage automatique pour mieux cibler les contrôles fiscaux sur la base des informations que les contribuables dévoilent eux-mêmes sur les réseaux sociaux. Le commerce illicite et les fausses domiciliations fiscales sont particulièrement visés par l'article 57 du projet de loi de finances 2020, qui prévoit un usage de l'intelligence artificielle au service de la lutte contre la fraude fiscale, adopté le 13 novembre dernier par les députés[3]. Ainsi, ce projet prévoit de renforcer les moyens informatiques pour améliorer le ciblage des opérations de contrôle fiscal grâce à un investissement de 20 millions d'euros d'ici à 2022.

L'intelligence artificielle, souvent invoquée pour discourir sur les technologies numériques, est un terme trompeur, car il évoque les capacités de machines fantasmées par les œuvres de science-fiction popularisées par le septième art. En matière fiscale, rien de tel : parmi la myriade des comportements observés, l'utilisation de techniques d'apprentissage automatique a pour objectif de détecter ceux récurrents qui seraient spécifiques à certains types de fraudes (à la TVA, au blanchiment, à la fausse domiciliation et à l'optimisation illicite).

Depuis la fin du secret bancaire en Suisse, le Service de traitement des déclarations rectificatives des repentis fiscaux a encaissé 8,3 Md€ en redressements, mais le flux se tarit désormais. Un guichet similaire doit s'ouvrir aux entreprises souhaitant revenir sur des pratiques douteuses contre une réduction de pénalités. Cependant, l'administration fiscale concède un point faible : *« Aujourd'hui, près d'une vérification sur quatre n'aboutit qu'à un redressement peu élevé. »* Ainsi, depuis le pic de

1Rapport n°1236 de la *Mission d'information relative à l'évasion fiscale internationale des entreprises*, présenté le 12/09/2018 par Mme Bénédicte Peyrol, 311 p.

2« L'intelligence artificielle est un élément pour mieux orienter les contrôles », *Le Figaro*, 07/11/2019.

3« Fisc, vers une surveillance des réseaux sociaux », A.Tonnelier, *Le Monde*, 15/11/1019.

2015 dû aux redressements exceptionnels de *Vivendi* (1,4 Md€) et de *Google* (1,1 Md€), le fisc n'arrive à recouvrer que 50 % environ des sommes réclamées. C'est pourquoi, la Cour des comptes note dans son dernier rapport que *« les résultats demeurent en retrait des ambitions affichées »*[4].

Selon Gérald Darmanin, l'algorithme permet *« juste d'établir les plans de contrôle les plus pertinents en ciblant mieux les personnes à contrôler »*. D'autre part, il s'agirait d'une expérimentation sur trois ans ne couvrant que trois types de fraudes : le trafic de marchandises prohibées, l'activité professionnelle non déclarée et la domiciliation fiscale frauduleuse. L'administration fiscale prévoit dans son projet de conserver les données utiles après filtrage sur une période de trente jours. Cependant, le champ des données prospectées par cette intelligence artificielle dans le cyberespace serait très étendu puisqu'il concernerait les réseaux sociaux comme *Facebook*, les messageries comme *Instagram* ou *Twitter* et les sites de commerce en ligne comme *LeBonCoin* ou e*Bay*. Les algorithmes d'apprentissage automatique ont été conçus par la Mission requêtes et valorisation, une cellule d'une vingtaine de spécialistes dédiée à la fouille des données. Ainsi, la fraude de certaines entreprises prétendant ne pas posséder d'établissement en France aurait été détectée à partir des curriculum vitae sur *LinkedIn* de leurs cadres habitant la région parisienne. Les particuliers ne seraient pas épargnés : des photos aériennes récupérées sur *Google Map* auraient également permis de détecter nombre de piscines privées non déclarées à la taxe foncière.

Dispersées selon plusieurs applications, les données concernant les 5 millions d'entreprises françaises imposables sont rapprochées des données publiques sur leur activité (par exemple, registre du commerce, dépôts de marques et de brevets) et des différentes informations concernant les 37 millions de foyers fiscaux, qu'il s'agisse de sources fiscales, bancaires, immobilières ou sociales. Les algorithmes de sélection se fondent sur une modélisation des indicateurs de fraude basés sur les profils des entreprises contrôlées depuis 2001 pour identifier les comportements suspects. Cependant, l'amélioration des performances du logiciel s'avère laborieuse, car elle reste tributaire de la qualité des données d'apprentissage. Néanmoins, l'efficacité des algorithmes d'apprentissage devrait bénéficier de l'entrée en vigueur depuis juillet 2018 de la *Convention multilatérale pour la mise en œuvre des mesures relatives aux conventions fiscales pour prévenir l'érosion de la base d'imposition et le transfert de bénéfices*[5] signée par plus de 100 pays dont le Panama et la Suisse. Réciproquement, toute entreprise dont le chiffre d'affaires dépasserait 750 millions d'euros est désormais tenue de transmettre au fisc français un rapport par pays de ses activités à l'international indiquant ses effectifs, bénéfices et impositions.

4Source: https://www.ccomptes.fr/sites/default/files/EzPublish/09-lutte-contre-fraude-fiscale-RPA2016-Tome-2.pdf, consultée le 25/11/2019.

5Source : https://www.oecd.org/fr/fiscalite/conventions/convention-multilaterale-pour-la-mise-en-oeuvre-des-mesures-relatives-aux-conventions-fiscales-pour-prevenir-le-BEPS.pdf, consultée le 25/11/2019.

Toutefois, la Commission nationale informatique et libertés et, à sa suite, un certain nombre de membres de la Commission des lois de l'Assemblée nationale se sont émus *« des conséquences sur les droits au respect de la vie privée et à la protection des données, mais aussi sur la libre communication des pensées et des opinions ».* C'est pourquoi les députés de la majorité souhaitent restreindre la durée de l'expérimentation à deux ans en la limitant aux seuls champs de l'économie souterraine et de la fausse domiciliation à l'étranger qui constituent des faits graves et difficiles à identifier, excluant celui de l'activité professionnelle non déclarée. Pour éviter une inflation de commentaires malveillants, voire de délations, ces députés veulent limiter l'exploitation aux seules informations manifestement publiées par les intéressés eux-mêmes. Enfin, ils prévoient par amendement de restreindre le délai de conservation des données à cinq jours et d'exclure le recours à la sous-traitance dans la chaîne des traitements numériques.

Plus radicale, la Quadrature du Net[6] a appelé le Parlement à *« rejeter le flicage fiscal des réseaux sociaux »*[7] en vain, du moins pour l'instant. Pour cette association, *« il s'agit ici de prendre un bazooka pour tuer une mouche »* : l'outil serait disproportionné par rapport à la liste des types de délits énumérés certes nombreux, mais dont l'impact est jugé mineur par rapport aux risques encourus sur d'éventuelles atteintes aux droits des personnes. Même si la lutte contre la fraude fiscale est reconnue par le Conseil constitutionnel comme objectif pouvant légitimer certaines restrictions en matière de libertés individuelles, l'article 57 du projet de loi serait contraire à la Constitution française et aux règlements de l'Union européenne (*ibid.*). En outre, cette association rejette la qualification d'expérimentation arguant que ce dispositif concernerait d'emblée l'ensemble du territoire français.

Pour leur part, les organisations syndicales ont exprimé la crainte que le recours aux algorithmes d'apprentissage automatique vienne accélérer les suppressions de postes en cours : depuis 2010, ce sont plus de 3 000 emplois qui ont été supprimés sur les 12 000 postes affectés au contrôle fiscal. En effet, la sanctuarisation des effectifs du contrôle fiscal ne s'applique qu'aux vérificateurs externes (4 500 emplois), mais pas aux emplois des autres services de l'instruction des dossiers, chargés du contrôle des pièces, de la recherche d'anomalies et de la vérification des critères de sélection. Selon les organisations syndicales, cette baisse des effectifs se paierait par une perte de qualité sur les données d'apprentissage et une efficience diminuée dans le traitement des dossiers délictueux.

6 Initialement constituée en 2008 comme une association de fait puis fondée en mars 2013 comme association sous le régime de la loi de 1901, la Quadrature du Net se donne pour but la défense des libertés sur Internet avec pour paradigme le maintien de la « neutralité du réseau ».

7Source : https://www.laquadrature.net/2019/11/05/le-parlement-doit-rejeter-le-flicage-fiscal-des-reseaux-sociaux/, consultée le 25/11/2019.

De facto, l'innovation financière et juridique des conseils en optimisation fiscale est une source constante d'inventivité frauduleuse : si les grandes entreprises multinationales focalisent désormais l'attention en matière d'évasion fiscale, les partenariats d'entreprise constituent aux USA une source de préoccupation croissante pour l'administration fiscale. Beaucoup moins susceptibles d'être audités, ils sont devenus les véhicules privilégiés de nouvelles stratégies de fraude. Très utilisés par les fonds de capital-risque ou de couverture de risque de marché, les sociétés immobilières et les groupes pétroliers ou gaziers, ces partenariats y sont actuellement mobilisés pour bénéficier d'une niche fiscale controversée, le *Installment-sale Bogus Optional Basis* (*iBOB*). En gonflant artificiellement la valeur de base d'un actif dans une déclaration de revenus afin d'éliminer les gains imposables lorsque cet actif est vendu, les transactions *iBOB,* bien que parfaitement légales individuellement, aboutissent à une déduction frauduleuse dans le résultat global pour certaines chaînes de transactions. Retranscrivant les lois fiscales fédérales (soit près de 75 000 pages de réglementation), puis codant les transactions douteuses et l'enchevêtrement des prises de participation au sein des partenariats, des chercheurs du *Massachusetts Institute of Technology* et de la *MITRE Corporation* ont mis en évidence les combinaisons spécifiques d'opérations et les structures de partenariat susceptibles d'aboutir à des chaînes frauduleuses de transactions *iBOB*[8]. De tels travaux pourraient inspirer opportunément les futures applications de lutte contre les fraudes en réseaux qui sont souvent la marque tutélaire d'organisations criminelles à caractère mafieux.

Au niveau français comme européen, l'éventualité de renforcer les contrôles dans certains domaines économiques est désormais clairement envisagée. En 2007, le Conseil des prélèvements obligatoires évaluait les pertes dues à la fraude pour le budget de l'État entre 20,5 et 25,5 Md€ par an, soit environ 8 % des recettes fiscales. En 2010 et 2013, les rapports de la Cour des comptes relatifs à la lutte contre la fraude fiscale et au contrôle fiscal mettaient en évidence la faiblesse du pilotage du dispositif français et « *l'inadaptation de son organisation aux risques de fraude complexes et à fort enjeu financier* ». En 2014, les 1,5 million de contrôles fiscaux réalisés ont donné lieu à la notification de 19,3 Md€ de redressements. Cependant, au cours de cette même année, seulement 10,4 Md€ de créances issues du contrôle fiscal ont été recouvrées. L'Office européen de lutte anti-fraude (Olaf) annonce en 2017 avoir clôturé 197 enquêtes, recommandant le recouvrement de 3 Md€ pour le budget de l'Union européenne, chiffre « *exceptionnellement élevé* » résultant de fraudes par sous-évaluation, clôturées par l'Olaf au cours de l'année. Dans le même temps, l'Olaf ouvre 215 nouvelles enquêtes, faisant suite à 1 111 analyses préliminaires effectuées par ses experts.

Compte tenu de ces moyens supplémentaires et des pouvoirs discrétionnaires d'investigation de ces services de lutte contre la fraude fiscale, il conviendrait d'en renforcer le contrôle par l'autorité

8 Source : Hemberg, E. *et al.* (2016) "Detecting tax evasion: A co-evolutionary approach". *Artificial Intelligence and Law*, 24, 2, pp. 149-182. https://doi.org/10.1007/s10506-016-9181-6

judiciaire. Or, il semble qu'on en prenne guère le chemin : ainsi, le magistrat responsable depuis août 2015 de l'agence *Tracfin*, cellule anti-blanchiment de Bercy, a été remplacé le 10 juillet 2019 par une inspectrice des finances.

Affaires à suivre, donc… *Dominique Desbois*[i]

iNDLR : depuis janvier 2019, Dominique Desbois est le représentant français au sein du *Technical Committee* « *ICT and Society* » (*TC9*) de l'*International Federation for Information Processing* (*IFIP*) pour une durée de trois ans.