

## 16ème législature

<b>Question N° :</b> <b>5364</b>	De <b>M. Aurélien Saintoul</b> ( La France insoumise - Nouvelle Union Populaire écologique et sociale - Hauts-de-Seine )	<b>Question écrite</b>
<b>Ministère interrogé</b> > Solidarités, autonomie et personnes handicapées		<b>Ministère attributaire</b> > Solidarités et familles
<b>Rubrique</b> > prestations familiales	<b>Tête d'analyse</b> > Contrôle des ayants droits de la CNAF à partir de leurs données personnelles	<b>Analyse</b> > Contrôle des ayants droits de la CNAF à partir de leurs données personnelles.
Question publiée au JO le : <b>07/02/2023</b> Réponse publiée au JO le : <b>17/10/2023</b> page : <b>9240</b> Date de changement d'attribution : <b>21/07/2023</b> Date de signalement : <b>06/06/2023</b>		

### Texte de la question

M. Aurélien Saintoul interroge M. le ministre des solidarités, de l'autonomie et des personnes handicapées à propos de l'utilisation de la technique d'exploration de données, ou datamining, par la Caisse nationale des allocations familiales. Cette technique prétend automatiser l'identification d'une erreur au sein d'une grande quantité de données. Depuis 2010, la CNAF utilise un algorithme de notation de ses allocataires fonctionnant à partir de données administratives les concernant. Cette technique est censée mesurer les risques d'erreurs ou les fraudes dans le dossier d'un allocataire. Plus ce « risque » est élevé, plus la probabilité que celui-ci soit contrôlé augmente. L'ancien directeur général de la CNAF déclarait en juin 2022 que l'exploration de données était devenue « la première source de détection des dossiers destinés au contrôle ». Néanmoins, les algorithmes sont défectueux et portent en eux les biais de ceux qui les codent. Dans ce cas précis, une enquête de *France Info* nous apprenait le 9 décembre 2022 que le fait de bénéficier de plusieurs prestations était de nature à fausser l'algorithme. Si les variations de ressources ou les changements dans une situation familiale sont des événements communs, ils peuvent tromper l'algorithme, entraîner un contrôle et même une suspension préventive des versements. Dès lors, cette situation suscite de nombreuses interrogations. Des journalistes et des associations ont tenté d'obtenir le code de l'algorithme utilisé par la CNAF. Selon *France Info*, le score de risque est d'emblée plus élevé pour les personnes les plus précaires et une circulaire interne de la CNAF datée de 2012 recommandait de « cibler les personnes nées hors de l'Union européenne ». Comment fonctionne cet algorithme ? Prend-il en compte des données subsidiaires ou discriminantes ? Malgré des erreurs manifestes et des facteurs de notation troubles, ce système place les contrôleurs dans la quasi-obligation de détecter des irrégularités dans les dossiers au score fortement élevé. En conséquence, il aimerait savoir comment M. le ministre contrôle cette pratique de la CNAF, s'il peut garantir qu'elle ne discrimine ni socialement ni ethniquement les ayants droits dans le pays, s'il est en mesure de fournir le code de l'algorithme dans un langage intelligible et s'il peut fournir une liste claire et complète des variables qui déterminent le score de risque d'un allocataire.

### Texte de la réponse

Le Gouvernement s'est fait communiquer les caractéristiques principales de l'algorithme utilisé par le réseau des



Caisses d'allocations familiales (CAF) dans leurs démarches de lutte contre les erreurs constitutives ou non d'une fraude. Ces démarches sont bien entendu nécessaires : avec plus de 95 milliards d'euros de prestations - versées à 13,6 millions de foyers représentant plus de 30 millions de personnes – conditionnées en partie par déclarations, les CAF se doivent d'assurer l'exactitude de celles-ci, ne serait-ce que pour l'application du principe du « paiement à bon droit ». A titre d'exemple illustrant cette nécessité, les éléments justificatifs de la prime d'activité connaissent un taux d'erreur avant contrôle de 50 %. Globalement, sur les 4 millions de contrôles réalisés en 2021, 740 000 ont permis de détecter des indus, dont parmi eux 43 000 de frauduleux, c'est-à-dire comportant une intention avérée de fraude. Parallèlement, 351 000 ont permis de générer des rappels de droits, en faveur de l'allocataire. La ministre des solidarités et des familles est en mesure de vous fournir quelques caractéristiques de l'algorithme - ou plus exactement de l'ensemble d'algorithmes regroupés sous l'appellation de « datamining » - permettant ce contrôle : - il porte pour l'essentiel sur les revenus déclarés et concerne donc à titre principal les prestations sous conditions de ressources (revenu de solidarité active, prime d'activité, allocations logement...) qui appellent des données déclaratives ; ces prestations ont des fréquences de calcul trimestrielles : ce sont donc aussi celles faisant l'objet des contrôles les plus fréquents et les plus nombreux ; - les techniques algorithmiques en jeu ont été mises au point depuis une dizaine d'années à partir de l'étude objective des situations générant potentiellement le plus d'erreurs au regard de la réglementation et visent uniquement à minimiser la probabilité de générer un indu ou un rappel de versement aux allocataires ; - seules les données détenues par les CAF sont utilisées, c'est-à-dire des informations utiles pour la vie du dossier de l'allocataire (éléments d'identification, situation familiale, professionnelle, données financières, résidence des allocataires, etc.) au regard des règles de droit qui encadrent le bénéfice des prestations (notamment les conditions de ressources et de résidence). Par ailleurs, la caisse nationale d'allocations familiales (CNAF) communique chaque année en toute transparence sur les principes de sa politique de contrôle et de lutte contre la fraude, dont l'usage du datamining. Elle communique également les données individuelles retenues la concernant à toute personne qui en fait la demande, dans le respect du Règlement général de la protection des données. Outre cette politique d'ouverture, la CNAF se doit d'être prudente quant à sa politique de divulgation, dans la mesure où il est d'intérêt général d'éviter la possibilité de stratégies de contournement conçues et susceptibles d'être mises en œuvre, faute de ces précautions, par des personnes mal intentionnées. De fait, l'utilisation de dispositifs de fraude de plus en plus sophistiqués (i.e. capables de contourner un grand nombre de mesures de contrôle) a malheureusement été constatée ces dernières années. En tout état de cause, l'objet de ces algorithmes est de cibler les dossiers en raison des probabilités d'erreurs, voire de fraudes. Ces échantillons conduisent à cibler les dossiers contrôlés par un agent et n'ont pas de conséquence automatique. Les contrôles sur pièces et sur place que les agents réalisent sont sans incidence sur les droits des allocataires s'ils se soumettent aux contrôles et que leur dossier ne présente pas d'anomalie. Il va de soi qu'une éventuelle anomalie peut faire apparaître un trop perçu mais aussi conduire à un rappel de droits au bénéfice de l'allocataire. Ainsi, ces préoccupations ne doivent pas occulter que les techniques de datamining pourront aussi être utilisées de façon croissante pour favoriser l'accès aux droits pour les 20 à 30 % de personnes qui n'y recourent pas. C'est ainsi qu'ont été réalisées des campagnes d'appels sortants ciblés sur des allocataires repérés au préalable, et qui ont permis d'ouvrir des milliers de droits à la prime d'activité ou à l'allocation de soutien familial. De nouvelles campagnes du même ordre sont organisées cette année, dans l'objectif de les reconduire au-delà de 2024 dans le cadre de la mise en œuvre de la « solidarité à la source ».