

Réunion du 4 novembre 2025 d'un groupe de travail du comité d'éthique des usages des données, des algorithmes et des codes sources de la Cnaf sur le datamining données entrantes (DMDE) 2025

DSER 20/11/2025

Participants :

Membres du comité d'éthique

Collège associatif représentant les usagers

Jean-Claude Boual

Collège d'experts

Caroline Lequesne

Bertrand Pailhès

Collège du Conseil d'administration de la CNAF

Jacques Buisson

Antoine Math

Collège de professionnels CNAF et Caf

Clémence Coulette représentant Estelle Louis (Caf 16)

Lucie Gonzalez CNAF-DSER

Arnaud Rozan CNAF-DES

Autres participants CNAF/CAF

Maxime Blanchot CNAF-MACSSI

Laurine Bois CNAF-MACSSI

Kevin Chilin CNAF-DSER

Thomas Desmoulins CNAF-DR

Stéphane Donné CNAF-DES

Claire Laporte CNAF-DSER

À partir de l'exemple des familles monoparentales et sur la base d'une note envoyée en amont de la réunion, ce groupe de travail a été organisé pour échanger, en visant une portée plus générale, sur les biais de nature discriminatoire du DMDE 2025, les mécanismes qui expliquent la sur-représentation dans les scores élevés, les points de vigilance que les membres du comité souhaitent mettre en avant et les indicateurs de suivi à mettre en place notamment une fois le modèle déployé (début 2026). Cette note reprend les principaux points de la discussion.

1. Contexte global

a. Modèle de ciblage et discrimination

- Il est rappelé que les enjeux soulevés par le DMDE de la Cnaf s'inscrivent dans un contexte européen de risque élevé de contentieux autour des modèles de ciblage. Le Conseil de l'Europe invite à les abandonner depuis 2017, ce qui implique de privilégier une sélection sur une base aléatoire ou, du moins, de justifier de façon très rigoureuse le recours au ciblage. Le débat sur des contrôles aléatoires sera donc difficile à éviter pour le DMDE de la Cnaf. Il ne faut pas minimiser le contexte de forte sensibilité et défiance vis-à-vis de ces modèles, renforcé en cas d'automatisation de la décision. La question de l'intervention et de la marge de manœuvre humaines (« garantie humaine ») dans l'utilisation de l'outil est centrale. Il y a un risque réel de requalification en décision automatisée même si une personne intervient dans le processus, s'il est considéré que le score est déterminant dans la prise de décision (cf. arrêt Schufa). Dans cette perspective il serait utile de valoriser les actions d'accompagnement et de formation des contrôleurs car elles contribuent à une meilleure maîtrise des effets du modèle lorsqu'il est mis en œuvre. A noter que dans l'usage du modèle 2018, le fait que 37% des dossiers avec les plus hauts scores (égaux ou supérieurs à 0,9) ne sont pas contrôlés dans les douze mois porte à croire que les décisions de contrôle ne sont pas automatiques. Cette évaluation sera utile à reproduire après quelques mois d'utilisation du DMDE 2025.
- C'est à la Cnaf de démontrer que ses modèles ne sont pas discriminatoires (charge de la preuve).

b. Communication autour du modèle

- Dans ce contexte, il est recommandé de communiquer largement les éléments de documentation du modèle de façon transparente, avec un effort de vulgarisation pour les rendre accessibles aux publics les moins initiés, et intelligible pour les personnes concernées (les allocataires) : objectifs, place dans la politique de contrôle de la Branche Famille, méthode de construction, contenu, effets du

modèle et mécanismes sous-jacents...pour limiter la suspicion et permettre un débat sur des bases informées.

- La documentation du modèle et de ses effets sera aussi utile pour objectiver les termes du débat autour de la fraude sociale et de la rationalisation de l'action publique.
- Cette communication sur le modèle devra faire ressortir le fait que les Caf contrôlent plus les personnes qu'elles accompagnent aussi davantage.
- Un point de vigilance a été soulevé concernant la communication des données sur les erreurs déclaratives des familles monoparentales présentes dans la documentation du DMDE. Il ne faudrait pas que cela ravive les controverses qui ont prévalu autour de l'Allocation de parent isolé (API) dans les années 1985 à 2010 : suspicion de travail dissimulé ou de fausses déclarations de situation d'isolement, désincitation au travail...

2. Analyse du modèle en regard des 26 critères de discrimination

a. Discrimination directe

- Les variables correspondantes ont presque toutes été écartées d'emblée lors de la construction du modèle.
- Deux points méritent des précisions :

Critère de Situation économique

Plusieurs variables du modèle liées au niveau de ressources ou à la perception des prestations reflètent la situation économique du foyer allocataire. Cela est inévitable puisqu'une large part des prestations versées sont conditionnées aux ressources, en particulier les prestations les plus risquées. La seule alternative serait une sélection aléatoire des dossiers, où chaque personne a la même probabilité d'être contrôlée, ce qui revient à renoncer au ciblage, et donc au modèle.

Critère de Situation familiale (familles monoparentales notamment)

Les éléments produits par la Cnaf (pages 9 à 13 de la note) conduisent à conclure à l'absence de biais de discrimination directe des familles monoparentales introduit par la variable *d'Activité du responsable de dossier et de son conjoint*. Cette variable vise principalement à capter le risque élevé associé aux situations des couples où le conjoint du responsable de dossier est déclaré inactif. Dans ce cas, il semblerait que les informations sur la situation professionnelle du conjoint (et le cas échéant sur ses revenus d'activité) soient souvent mal renseignées ou imparfaitement mises à jour, alors que les données déclarées seraient plus fiables pour les couples monoactifs où c'est le conjoint qui travaille.

Cependant le découpage de cette variable en quatre modalités doit être explicité pour lever toute ambiguïté sur un risque de discrimination directe des familles monoparentales. Il repose en effet, de façon jugée peu intuitive, sur les différentiels de risques d'indus longs et pas sur la notion socioéconomique d'activité. De fait, les familles monoparentales sont présentes à 98% dans une même modalité « Autres situations », sans distinction entre celles ayant un emploi et les autres, et mêlées à d'autres populations (personnes seules non retraitées avec ou sans activité et avec ou sans enfant ; couples monoactifs où c'est le conjoint du responsable de dossier qui travaille). Cette modalité « Autres Situations » est majoritaire (68%). Comme elle est prise comme modalité de référence, il n'y a aucun sur- ou sous-risque associé (l'odd-ratio est égal à 1). Il est recommandé d'expliciter ce découpage, ou de le modifier en faveur d'une logique socioéconomique, car l'incompréhension pourrait susciter de la suspicion.

b. Discrimination indirecte

- Le graphique 7 de la note est central pour appréhender la question de la discrimination indirecte. Il doit être complété de plusieurs chiffres : le poids au sein des indus longs et importants (27% pour les familles monoparentales, cité p.7 de la note) ; les poids dans les 1% et 2% des scores les plus élevés issus du DMDE 2018 et des DMDE 2025. Pour les 1% de scores les plus élevés issus du DMDE 2018, ce poids est de 49% pour les familles monoparentales. En effet, il est précisé par la Cnaf que la très grande majorité des dossiers contrôlés sont dans les 5% des scores de risque les plus élevés. Cependant, l'inverse n'est pas vrai : du fait du classement des dossiers par score décroissant dans les outils à destination des contrôleurs, les dossiers dont les scores sont dans les 1% les plus élevés sont bien plus contrôlés (quasi exhaustivement) que les dossiers figurant plus bas dans la liste. A noter que la part des familles monoparentales dans les 5% de scores les plus élevés du modèle après pré-remplissage est plus basse (40%) que dans le modèle portant sur la période antérieure à la mise en œuvre de la solidarité à la source (43% pour le modèle 2018 et le modèle 2025 avant préremplissage).
- A ce titre, l'information sur les volumes de contrôles au titre du datamining, actuel et prévu, apparait doublement essentielle. Elle permet d'une part de relativiser la portée de l'outil et donc ses effets (seulement 160 000 contrôles DMDE en 2024). D'autre part, plus le nombre de contrôles datamining visé sera important, plus on « descendra bas » dans la liste des scores et moins les effets seront concentrés sur les populations les plus à risque. Autrement dit, compte tenu du risque d'indu mesuré pour les familles monoparentales (31% de dossiers avec un indu long, contre 18% en moyenne pour l'ensemble de la population allocataire), on peut s'attendre à un poids des familles monoparentales de l'ordre de 50% si on contrôle 1% des 13,8 millions de foyers allocataires (soit 138 000 dossiers), et autour de 40-43% si on en contrôlait 5

fois plus (5% de scores les plus élevés, soit 690 000 dossiers). Compte tenu de la disparition de nombreux contrôles devenus obsolètes avec la solidarité à la source, les capacités de contrôle des Caf devraient être redéployées vers les contrôles datamining et la lutte contre la fraude à enjeux (fraudes à résidence, déclaration des ressources issues de l'économie des plateformes...). En outre, la répétition des contrôles sera mieux maîtrisée. Cela devrait donc conduire à contrôler largement plus que les 1% des scores les plus élevés.

- A cet égard, une mise à jour de la note sur la politique de contrôle présentée en juin au comité d'éthique sera bienvenue, complétée par la volumétrie des différents contrôles et en particulier ceux générés par le DMDE, en distinguant entre « contrôles sur pièces » et « contrôles sur place ».
- Le travail réalisé sur la situation familiale devrait être répliqué pour deux autres critères de discrimination : la nationalité et le handicap (via la perception de prestations spécifiques). L'analyse d'impact sur la protection des données (AIPD) intègre bien la question de la maîtrise des risques discriminatoires.
- Il n'est pas possible de mesurer la sensibilité des effets du modèle liés au choix de la cible (quels impacts, par exemple, d'une cible d'indus d'au moins 1000 € ?), au-delà des statistiques descriptives produites en amont et présentées à la séance de juin, qui n'ont pas fait ressortir de seuil plus pertinent, conduisant la Cnaf au *statu quo* (sachant que le seuil de 600 € représente moins en valeur réelle en 2025 qu'en 2018). En effet, il faudrait alors construire un modèle complet et refaire en particulier toute la phase de sélection des variables, ce qui représente une charge de travail disproportionnée.
- L'évaluation complète des effets du modèle nécessite de disposer, en plus des statistiques fournies sur le modèle « théorique », des données observées « en vie réelle », une fois le modèle déployé dans les Caf, pour tenir compte de l'impact de la façon dont le modèle est utilisé par les contrôleurs. Il paraît en effet difficile de faire l'hypothèse que les écarts de sur- ou sous-représentation au sein des 5% (ou 1%) des scores les plus élevés du modèle et celle au sein des contrôles effectivement réalisés sur la base du DMDE soient les mêmes pour le modèle 2025 que pour le modèle 2018, compte tenu :
 - du bouleversement que constitue la solidarité à la source du paysage des erreurs déclaratives et, conséquemment, de la politique de contrôle. Les contrôles portant par nature sur le passé, avec un recul de deux ans (hors fraude), les effets de la solidarité à la source sur les actions de contrôles ne seront pleinement montés en charge qu'en mars 2027. Cette réforme constitue en outre une opportunité pour la Branche de mieux articuler les contrôles et limiter la

réitération. Il est en particulier prévu d'allonger la durée minimale entre deux contrôles, de 12 ou 18 aujourd'hui à 24 mois à partir de mars 2026 ;

- de la méthodologie plus soignée d'intégrer la dimension éthique adoptée pour construire le modèle 2025 ;
- du fait que le modèle 2025 a été élaboré à partir de données antérieures à la réforme ; avec ses deux volets, applicables avant et après pré-remplissage, c'est un modèle de transition.