

Séance du comité d'éthique sur les usages des données, des algorithmes et de l'intelligence artificielle de la Cnaf du 6 octobre 2025

Membres du comité participants :

1. Collège experts
 - ❖ Bruno Bachimont, professeur de philosophie et de *computer science* à l'Université de technologie de Compiègne
 - ❖ Elise Debiès, avocate en droit de la protection des données, enseignante à la Sorbonne et à l'IMT Paris, membre du Conseil d'orientation du CRAPS
 - ❖ Bertrand Pailhès : ingénieur en chef des mines, directeur des opérations et des territoires à l'Institut national de l'information géographique et forestière (IGN)

2. Collège associatif
 - ❖ Jean-Claude Boual, collectif Changer de Cap
 - ❖ Maryse Artiguelong, la Ligue des Droits de l'Homme, collectif Alerte
 - ❖ Jean-Paul Thierry, Secours Catholique, collectif Alerte
 - ❖ Henri Simorre, ATD Quart Monde, collectif Pacte du pouvoir de vivre
 - ❖ Nathalie Latour, Fédération des acteurs de la Solidarité, collectif Pacte du pouvoir de vivre

3. Collège des représentants du Conseil d'administration
 - ❖ Présidente : Isabelle Sancerni
 - ❖ Représentant des organisations de salariés : Franck André (CFDT)
 - ❖ Représentant des organisations d'employeurs : Béatrice Gruau Racine (MEDEF)
 - ❖ Représentant des associations familiales : Jacques Buisson (UNAF)
 - ❖ Personnalité qualifiée : Antoine Math (Économiste, chercheur à l'Ires et collaborateur scientifique auprès du HCFEA)

4. Collège des professionnels de la Cnaf et des Caf
 - ❖ Christelle Dubos, Médiatrice nationale et directrice de la protection des droits, des valeurs de la République et de la déontologie
 - ❖ Lucie Gonzalez, Directrice des statistiques, des études et de la recherche
 - ❖ Estelle Louis, Directrice de la Caf de la Charente
 - ❖ Jérôme Roteta, Directeur de la Caf des Pyrénées-Atlantiques

- ❖ Arnaud Rozan, directeur général délégué, chargé du pilotage stratégique et de la coordination des évaluations

- ❖ Excusés : Caroline Lequesne : Maîtresse de conférences en droit public à l'Université Côte d'Azur (UCA), Directrice du Master II Droit algorithmique et gouvernance des données ; Christian Gelis, vice-président du Conseil d'administration de la Cnaf

Ordre du jour :

▪ **Approbation du compte rendu de la réunion précédente**

Le compte-rendu de la séance du 17 juin est approuvé à l'unanimité, moins une abstention. Il est convenu que l'éventuelle diversité des points de vue de chacun apparaisse dans les comptes rendus à venir des séances de cette instance délibérative.

▪ **Planification des prochaines séances**

La prochaine séance se tiendra le lundi 1^{er} décembre à 14 h. Les deux suivantes auront lieu en mars et juin 2026, les dates seront communiquées très prochainement.

▪ **Le troisième point de l'ordre du jour porte sur le modèle de *Datamining* « données entrantes » 2025.**

Dans un premier temps, la Cnaf propose un point de situation sur les tests du modèle de *datamining* présenté à la séance du 17 juin.

Pour rappel, ce modèle ne prend pas en compte les effets du pré-remplissage des déclarations trimestrielles de ressources (DTR) dans le cadre de la réforme de la solidarité à la source. Il comporte 17 variables, regroupées pour la présentation selon quatre catégories : les informations sur le foyer allocataire ; les ressources ; les prestations et les changements dans les dossiers.

Les tests, réalisés entre avril et juillet, ont permis d'évaluer les performances de ce modèle 2025 et de les comparer avec le *datamining* appliqué actuellement (modèle 2018). Il en ressort que le modèle 2025 présente de meilleures performances, tant pour la qualité du ciblage (pourcentage de dossiers testés présentant des indus d'au moins 600 € pendant 6 mois) que pour les montants d'indus moyens par dossier testé.

La Cnaf présente ensuite la construction d'un deuxième modèle qui prend en compte les effets du préremplissage des déclarations trimestrielles de ressources (DTR).

La solidarité à la source est entrée en vigueur en mars 2025 pour le RSA et la prime d'activité. Les déclarations trimestrielles de ressources (DTR) sont désormais préremplies avec les ressources connues à partir du dispositif de ressources mensuelles (DRM). Cette réforme devrait donc sensiblement réduire le risque d'erreur déclarative. Toutefois, l'allocataire dispose de la possibilité de signaler une anomalie éventuelle dans ce préremplissage. Dans les faits, ces « signalements » sont très peu fréquents et le plus souvent erronés. Lorsqu'un foyer allocataire signale un écart à la baisse dans au moins deux DTR portant sur un montant de 150 € chacune, le risque de figurer dans la « cible » du *datamining* est multiplié par 2,9.

Un second modèle a donc été construit pour améliorer le ciblage à partir de mars 2025. Les deux modèles, avec et sans prise en compte du préremplissage, coexisteront de janvier 2026, date de mise en service envisagée, jusqu'à mars 2027, date à laquelle deux ans se seront écoulés depuis la réforme (on remonte sur deux ans pour les indus et rappels). Pendant cette période, le score de risque sera calculé mensuellement à partir des deux modèles, chacun étant pris en compte au *pro rata temporis*. Ainsi, au fil du temps, le modèle intégrant le préremplissage contribuera de plus en plus au calcul du score de risque, jusqu'à le déterminer exclusivement à partir de mars 2027.

Ce second modèle comporte 11 variables, dont un grand nombre est commun au modèle ne prenant pas en compte le préremplissage.

Cette présentation a amené les participants à formuler les remarques suivantes.

- L'effort de transparence sur le modèle a été salué avec l'envoi d'une abondante documentation aux membres du comité.
- Il serait utile de faire davantage apparaître les différences entre les variables des deux modèles (présentées en annexe dans les documents envoyés avant la séance), notamment avec un tableau contenant l'ensemble des variables des modèles. Une note de lecture sera ajoutée pour faciliter la lecture des tableaux. Par ailleurs, la méthodologie consistant à écarter d'emblée certaines variables sensibles (par exemple, le genre ou la nationalité) sera rappelée, car elle était identique dans les deux modèles. Cela permet d'étayer l'absence de discrimination directe vis-à-vis de certains publics.
- Dans le prolongement des échanges de la séance précédente, une question a été posée sur l'asymétrie du modèle, ciblé sur les indus, tandis que l'erreur est symétrique et peut

consister aussi bien en indus qu'en rappels. La Cnaf précise qu'il existe une asymétrie dans la correction de ces erreurs dans le temps, par les allocataires eux-mêmes ou par les Caf : à 90 % pour les rappels, contre 40 % pour les indus. C'est ce qui explique le choix de cibler les indus dans le modèle. Par ailleurs, suite aux échanges de la séance précédente, certains tableaux ont été complétés avec des informations sur les rappels, car le contrôle est global et peut conduire aussi bien à un indu qu'à un rappel.

- L'affirmation selon laquelle, avec le préremplissage il n'y a plus d'erreurs DRM, est discutée. La Cnaf apporte quelques précisions. Le cas dominant est celui de la fiabilité des données du DRM. L'élément central est la confiance envers ces informations et la confiance à l'égard des allocataires qui ont la possibilité de les modifier. Malgré tout, le DRM peut parfois comporter des erreurs, le plus souvent des informations qui arrivent avec retard. Nous pouvons espérer que ces cas sont relativement peu nombreux, qu'ils pourront être corrigés à partir d'un signalement des allocataires ou, après coup, si les employeurs eux-mêmes rectifient les informations.

Dans un troisième temps, la Cnaf analyse les effets du modèle sur certains profils (*persona*) : bilan des deux modèles (cf. annexe sur les *persona*)

Un ensemble de statistiques descriptives est proposé pour certains profils d'allocataires. Les proportions parmi les 5 % dont les scores sont les plus élevés dans les deux modèles 2025 sont comparées avec celles au sein de l'ensemble de la population allocataire des Caf. Par ailleurs, trois situations alternatives sont examinées : le modèle actuellement utilisé (modèle 2018), les contrôles qui ont été lancés en 2024 à partir de ce *datamining* ; une sélection qui serait effectuée parmi les 5 % de dossiers ayant les montants de prestations les plus élevés.

Cette présentation a amené les participants à formuler les remarques suivantes.

- En réponse à une question, la Cnaf précise que le terme de « surreprésentation » ne correspond pas à un jugement de valeur, mais à un constat statistique.

- La comparaison avec le modèle actuel (2018) ne montre pas de grandes différences pour la plupart des prestations et des populations étudiées. La Cnaf confirme ce constat. Il peut s'expliquer par le fait que l'architecture générale des prestations n'est pas fondamentalement différente entre 2018 et 2025. Les modèles 2025 intègrent tout de même certaines évolutions

intervenues au cours des dernières années, notamment la réforme de la prime d'activité de 2019.

- Les informations fournies montrent que la Cnaf a pris de nombreuses précautions pour écarter le risque de discrimination directe vis-à-vis de telle ou telle catégorie. En revanche, il faudrait examiner s'il peut exister des cas de discrimination indirecte, notamment vis-à-vis des familles monoparentales. La Cnaf répond que les variables du modèle ne sont pas ciblées sur des catégories socio-démographiques en tant que telles. Elles dépendent plutôt des paramètres des prestations, certains conduisant à un risque plus important d'erreur que d'autres.

- La surreprésentation ne signifie-t-elle pas un surcontrôle concentré sur certaines catégories d'allocataires, sachant que 30 % des allocataires perçoivent 70 % des prestations ? Certains représentants du milieu associatif font part de témoignages qui leur sont remontés, notamment de la part de familles monoparentales contrôlées qui sont très dépendantes des prestations versées.

- L'équité de l'algorithme ne devrait pas seulement être examinée pour les profils surreprésentés, mais aussi en considérant ceux qui pourraient éviter tout contrôle (du moins au titre du *datamining*) du fait d'un niveau inférieur de score de risque. Il est aussi souligné qu'un algorithme ne peut satisfaire simultanément l'ensemble des critères d'équité que l'on peut se donner : un choix est nécessaire.

- Puisque le risque est lié à la difficulté pour les allocataires de remplir leurs dossiers, se pose le problème de la littéracie, fortement dissymétrique entre l'administration et les allocataires. L'enjeu consisterait donc à bien accompagner les allocataires pour fournir les informations. Ce point est souligné par les représentants du monde associatif ; certaines catégories sont surreprésentées, parce qu'elles cumulent plusieurs types de difficulté. L'enjeu de l'accompagnement est donc essentiel, tout comme celui de l'évolution du cadre légal (simplification). Les éléments fournis à partir du *datamining* permettent de distinguer les profils à risque en termes d'erreur, et donc de stabilité des parcours de vie. Cela pose des problèmes en termes de perception de l'équité et du sentiment d'être accompagné. Le sujet déborde le cadre du comité, mais il est majeur.

La Cnaf partage ce point de vue, l'enjeu de l'accompagnement et de la facilitation constitue un sujet essentiel, d'ailleurs bien au-delà du seul sujet du *datamining*. La réforme de 2025 répond en partie à cet enjeu grâce au préremplissage. Par ailleurs, le risque est plus élevé pour la prime d'activité que pour le RSA : il ne dépend donc pas seulement de la situation des

personnes, mais aussi des caractéristiques des prestations et des variations de ressources. Les personnes les plus éloignées de l'emploi, pour lesquelles le sujet de littéracie se pose sans doute davantage, bénéficient souvent du RSA sans ressources, ce qui est peu risqué du point de vue des indus de données entrantes. Mais le sujet de l'accompagnement, de la facilitation et de la simplification de certaines prestations constitue en effet un enjeu majeur.

- Il est souligné qu'un effort de pédagogie sera nécessaire. Certaines populations sont surreprésentées, non parce qu'elles sont directement ciblées, mais parce qu'elles se situent à l'intersection de plusieurs types de complexités. Les éléments présentés montrent qu'il y a des personnes à risque en termes d'erreurs et de stabilité de parcours de vie. L'amélioration de l'accompagnement, et plus largement de la législation constituent des enjeux de fond.

- Une question est posée sur l'écart entre les résultats théoriques issus du modèle et les contrôles effectifs. La Cnaf indique que ce point pourra être documenté ultérieurement pour les modèles 2025, au fur et à mesure que nous disposerons de suffisamment de données sur ce point, une fois les nouveaux modèles mis en production.

Le quatrième point porte sur la documentation d'une méthodologie de réduction des biais, à partir du profil des familles monoparentales

La Cnaf partage un document sur le sujet de la réduction des biais. L'objectif est de mettre en évidence les sur- ou sous-représentations statistiques du fait du *datamining*, d'en proposer des explications, ainsi que les méthodes ou actions mises en œuvre pour atténuer ce phénomène. Cette « mitigation » pourrait être effectuée *a priori*, au moment de la construction du modèle, ou *a posteriori* après que l'outil sera entré en fonctionnement.

Le document présenté porte sur les familles monoparentales : part des familles monoparentales parmi l'ensemble des allocataires (en nombre de dossiers et en montants de prestations versées), actions de la Branche famille envers ce public (accompagnement social, ARIPA, etc.), description des caractéristiques des prestations perçues (de manière à expliquer en quoi elles seraient susceptibles d'augmenter le risque d'erreur), situation des familles monoparentales vis-à-vis du risque d'indu d'au moins 600 € et pendant au moins 6 mois, actions menées pour limiter les biais du modèle, modalités d'utilisation du modèle dans les Caf...

Cette présentation a amené les participants à formuler les remarques suivantes.

- La fiche présentée montre que le risque d'indu est fortement majoré pour les familles monoparentales. Comment ce surrisque baisserait si les mesures d'accompagnement, « d'aller vers » et de soutien étaient renforcées ? La Cnaf indique que plusieurs types d'actions structurelles sont menées : solidarité à la source, rendez-vous des droits, accompagnement par les Caf pour remplir les déclarations. Par ailleurs, le contrôle ne doit pas être vu uniquement comme une sanction, mais comme un examen global de la situation de l'allocataire. Cela permet d'accompagner les personnes, notamment pour les aider comprendre les causes d'erreurs déclaratives.

- Il serait utile de comparer les effets d'une campagne de prévention des indus avec un contrôle ciblé. Il serait également utile de connaître précisément les causes des indus et des erreurs. La Cnaf indique qu'un travail d'analyse des causes des indus est en cours.

- Il serait utile de suivre dans le cadre du comité la montée en charge du préremplissage pour s'assurer que certaines sources d'erreur vont bien disparaître. Il serait également utile de bien distinguer les caractéristiques des prestations et des personnes qui en bénéficient.

- Il est demandé le volume des contrôles réalisés à partir du *datamining*. La Cnaf indique que les contrôles issus du *datamining* sont très minoritaires au sein de l'ensemble (5 % des contrôles sur pièce) et leur proportion diminue au fil du temps. Mais l'outil de ciblage reste important pour évoluer vers des contrôles à forte valeur ajoutée, avec une supervision humaine. Le *datamining* évite peut-être une sélection des dossiers reposant uniquement sur les critères de choix des contrôleurs, potentiellement biaisés.

- La solidarité à la source comporte une asymétrie d'information entre les allocataires et les services : l'allocataire doit bien comprendre les informations fournies et les utiliser. Du point de vue éthique, il ne faudrait pas que l'outillage qui est mis en place renforce cette dissymétrie informationnelle, sachant qu'il y a beaucoup d'éléments que l'allocataire ne maîtrise pas bien. La Cnaf répond que cette notion de dissymétrie peut être appréciée de manière inverse : les allocataires possèdent des informations sur leur situation que les Caf ne connaissent pas. Le travail des caisses consiste à acquérir ces informations. La solidarité à la source nous aide en ce sens, mais une partie de la dissymétrie demeure, notamment concernant la conjugalité.

- Il serait souhaitable de poursuivre ces débats sur le sujet du non-recours en exploitant une partie des éléments présentés à cette séance. Cela constituerait le contrepoint essentiel à la détection des indus. La Cnaf indique qu'il était bien prévu de consacrer une partie de la séance du 1^{er} décembre au versant de l'accès aux droits.

En conclusion de ce point, la Cnaf propose de poursuivre les travaux pour documenter les biais en réunissant un sous-groupe de membres du comité qui souhaitent y participer. Des dates seront proposées pour une réunion en mode distanciel. Les travaux du sous-groupe seront restitués en plénière à la prochaine séance du comité.

En conclusion, Nicolas Grivel remercie les membres du comité pour leur participation à cette troisième réunion. La prochaine séance permettra d'ouvrir une séquence sur l'accès aux droits, avec une mise en contexte et une présentation d'outils. Les séances de mars et juin permettront d'échanger sur l'actualisation de la charte éthique, de manière à prendre en compte des remarques transmises par certains membres du comité. Ces séances permettront également d'ouvrir le sujet de l'utilisation de l'intelligence artificielle dans la Branche. Elle n'est pas industrielle à ce stade, les questions éthiques peuvent varier suivant les sujets, mais il est important de débiter les échanges dans le cadre du comité.

Il est proposé de se retrouver le **1^{er} décembre après-midi, à 14h.**