De la transparence à l'explicabilité automatisée des algorithmes : comprendre les obstacles informatiques, juridiques et organisationnels

Denis Merigoux, Marie Alauzen, Justine Banuls, Louis Gesbert, Émile Rolley

De la transparence à l'explicabilité automatisée des algorithmes : comprendre les obstacles informatiques, juridiques et organisationnels

Denis Merigoux, Marie Alauzen, Justine Banuls, Louis Gesbert, Émile Rolley

RAPPORT DE RECHERCHE

N° 9535

2024-01

Équipes-Projet PROSECCO, LAMSADE

ISSN 0249-6399 ISRN INRIA/RR--9535--FR+ENG

De la transparence à l'explicabilité automatisée des algorithmes :

comprendre les obstacles informatiques, juridiques et organisationnels

Denis Merigoux*, Marie Alauzen†§, Justine Banuls*§, Louis Gesbert*§,
Émile Rolley‡§

Équipes-Projets PROSECCO, LAMSADE

Rapport de recherche n° 9535 — 2024-01 — 68 pages

Résumé : Les algorithmes publics ou traitements informatiques opérés par les administrations sont soumis à des obligations de transparence et d'explicabilité. Ces obligations ont été mises en place pour justifier les décisions prises par ces algorithmes et garantir leur redevabilité vis-à-vis des personnes concernées. Ce rapport de recherche vise à explorer les obstacles informatiques, juridiques et organisationnels, qui entravent la mise en œuvre de la redevabilité et propose d'élargir la conception de l'explicabilité, afin de la rendre plus opérationnalisable par les administrations. Partant d'un état de l'art de l'explicabilité des décisions automatisées ajusté aux spécificités administratives, nous avons initié une enquête exploratoire sur l'algorithme de calcul des aides au logement de la Caisse nationale d'allocations familiales et conçu, à partir de ce diagnostic, trois prototypes testant la faisabilité de la production automatique d'explications de qualité. Nous montrons ainsi que l'utilité des explications est tout aussi cruciale d'un point de vue externe – pour les citoyennes et les institutions de contrôle – que d'un point de vue interne à l'administration – afin d'assurer la fiabilité du système d'information au gré des évolutions. Nous recommandons que la technologie de la décision automatisée et celle de son explication partagent une infrastructure commune, s'appuyant sur la lettre du droit.

Mots-clés : algorithmes publics, droit, décisions automatiques, explicabilité, redevabilité

* PROSECCO, Inria Paris
† Chargée de recherche CNRS, Laboratoire d'Analyse et de Modélisation des Systèmes d'Aide à la Décision (LAMSADE)
‡ Développeur freelance, financé par la Direction Interministérielle du Numérique (DINUM)
§ Ces quatre auteurs ont contribué également à ce rapport. Marie Alauzen a contribué à la réflexion et à l'écriture, Justine Banuls, Louis Gesbert et Émile Rolley ont contribué à la réflexion et à l'élaboration des prototypes.

CENTRE DE RECHERCHE DE PARIS

2 rue Simone Iff - CS 42112
75589 Paris Cedex 12

From algorithmic transparency to automatized explanainability : understanding IT, legal and organizational challenges

Abstract : Public-sector algorithms or data processing carried out by public administrations are subject to obligations of transparency and explainability. These obligations have been put in place to justify the decisions taken by these algorithms and guarantee their accountability to the people concerned. The aim of this research report is to explore the legal and organisational obstacles to the implementation of accountability, and proposes that the concept of explicability be broadened to make it more operational for administrations. Starting with a state of the art of the explicability of automated decisions adjusted to the specific administrative specificities, we carried out an exploratory study of the algorithm for computing housing benefit at the Caisse nationale d'allocations familiales and, on the basis of this diagnosis, produced three prototypes testing the feasibility of automatically producing quality explanations. We have shown that the usefulness of the explanations is just as crucial from an external point of view –for citizens and institutions control– than from a point of view internal to the administration –in order to ensure the reliability of the information system at the end. We recommend that the technology underlying automated decision and that of its explanation share a common infrastructure, relying on the letter of the law.

Key-words : accountability, public algorithms, law, automated decisions, explanainability

Synthèse

L'obligation de motivation des décisions administratives fait partie des garde-fous de l'État de droit. Elle se manifeste par la production d'une explication de la décision prise par l'administration, à destination de la personne concernée. Dans la pratique, cette motivation peut recouvrir une pluralité de formes et se manifester par une simple interaction de guichet. Au fil de la modernisation administrative, l'informatisation des procédures et la centralisation d'opérations administratives devenues de plus en plus complexes ont introduit un nouveau mode de production des décisions, supposant la mise en place d'algorithmes permettant à l'administration de gagner en efficacité. Or, l'automatisation a transformé les enjeux de la motivation, amenant des problématiques d'informatique et de libertés et renouvelant l'exigence en direction de l'explicabilité d'algorithmes.

Or, l'explicabilité des algorithmes publics n'a rien d'une question purement technique devant restée confinée aux cercles experts de l'informatique. En France, Affelnet, ParcourSup et MonMaster, les algorithmes de répartition des élèves et des étudiants, ont alimenté de vifs débats sur les inégalités, le développement de stratégies et la capacité de l'État à gérer une telle infrastructure. Si l'ouverture des codes sources semble déjà un supplément de transparence, la mise à disposition sur internet des documents aussi techniques est loin d'avoir réglé les craintes des citoyens quant à la manière dont ils sont administrés. D'autant plus que, derrière ces cas sonores, qu'il existe une multitude d'algorithmes régissant la vie publique : les algorithmes de tarification déterminant l'éligibilité à la gratuité ou une réduction des droits d'accès aux services publics en matière de transports urbains, d'accès à l'eau ou de logement social ; les algorithmes de détection automatisée des infractions routières (stationnement illégal, franchissement de feux) ; les algorithmes de calcul des salaires des agents publics, des impôts, des amendes, des aides sociales, des droits à la retraite, etc. En d'autres termes, la vie publique est profondément structurée par des algorithmes, dont le public n'a souvent connaissance que lorsqu'ils dysfonctionnent.

Le droit de l'informatique et des libertés fixe un cadre pour la conception et l'usage des algorithmes des administrations. Le deuxième alinéa de l'article 47 de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés dispose que « pour ces décisions, le responsable de traitement s'assure de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard ». Cet article fixe l'horizon d'attente de l'explication des décisions administratives automatisées : elle doit être individualisée, détaillée et intelligible par la personne concernée.

Cependant, dans la pratique, toutes les décisions prises sur le fondement d'un traitement automatisé ne font pas l'objet d'explication, et lorsque le responsable du traitement fournit des explications, elles se caractérisent par leur brièveté, voire un certain hermétisme. Par exemple, les lignes du bulletin de salaire, bien que relativement détaillé et individualisé, restent indéchiffrables pour qui ne connaît pas déjà les mécanismes de calcul des cotisations sociales. Après avoir réalisé un état des lieux des pratiques administratives contemporaines, nous estimons qu'aucune forme explicative des décisions automatisée actuellement déployée ne remplit les exigences de la loi informatique et libertés.

Ce rapport de recherche propose d'élargir la conception de l'explicabilité des algorithmes publics, afin de la rendre plus opérationnalisable par les administrations. Nous affirmons que l'explicabilité des algorithmes publics n'est pas seulement un enjeu externe (de redevabilité vis-à-vis des personnes physiques et morales impliquées par ces traitements administratifs,

ni même vis-à-vis de la société politique dans son ensemble), mais qu'il s'agit, dans le même temps, d'un enjeu interne à l'administration.

Pour contribuer à la réflexion et renouveler les termes du débat, à partir d'une enquête sur l'explication de l'algorithme calculant les aides au logement, nous proposons que le système de décision automatisé soit employé pour déplier et expliquer toutes les étapes de l'algorithme. À partir de travaux menés à l'Inria dans le cadre du projet Catala, une méthodologie de programmation qui relie le code source du système aux textes qui fondent sa décision automatisée, nous explorons la faisabilité d'une production automatique d'explications individualisées, détaillées et intelligibles par les parties prenantes du contrôle interne et externe du système de décision. Nous avons élaboré trois prototypes pour le calcul des aides au logement.

Finalement, les résultats de cette enquête nous poussent à formuler une double recommandation concernant l'explicabilité des algorithmes publics. Premièrement, le système technique d'explication de la décision doit être aligné sur le système technique de prise de cette décision. Deuxièmement, le système de prise de décision doit être adossé sur les éléments qui fondent et justifient le traitement, et en particulier le droit qui spécifie, de manière plus ou moins détaillée, la plupart des décisions prises quotidiennement et massivement par les administrations. Nous pensons que c'est à ces deux conditions que l'horizon fixé par de la loi informatique et libertés pourra être approché.

Sommaire

1 Introduction : de la publication à l'explicabilité des algorithmes pour les administrations et les publics

En France, la création de ParcourSup en 2018, l'algorithme de répartition des élèves en première année de l'enseignement supérieur, a donné l'impulsion à une réflexion inédite sur la transparence des algorithmes. À partir du travail pionnier réalisé pour la publication, puis l'explicitation du fonctionnement de cet algorithme et de l'analyse des bonnes pratiques internationales, en 2019, la mission Etalab de la Direction interministérielle du numérique (DINUM) a édité un guide intitulé « Expliquer les algorithmes publics » [Pénicaud et Chignard 2019]. Partant de l'idée suivant laquelle la « redevabilité » de l'État vis-à-vis des citoyennes n'est pas soluble dans la publication en source ouverte, soit le fait qu'il ne suffit nullement de mettre en ligne le code source d'un programme informatique pour être transparent, mais que les administrations sont redevables des traitements automatiques dont elles usent au quotidien pour réaliser des missions de services publics, le guide définit les enjeux de la manière suivante :

> « Par rapport aux algorithmes mis œuvre par le secteur privé, les algorithmes publics ont des caractéristiques particulières :
>
> — Ils sont censés opérer au service de l'intérêt général,
> — Ils servent souvent à exécuter le droit, à (faire) appliquer la loi,
> — Ils sont bien souvent incontournables, c'est à dire qu'il n'existe pas d'alternatives pour les usagers.
>
> En ce sens, les algorithmes publics sont des formes de l'action publique et sont à ce titre soumis à la même forme d'exigence de redevabilité. Les administrations qui conçoivent et utilisent des algorithmes publics doivent donc "rendre des comptes" de leur utilisation auprès des individus concernés, mais aussi de la société dans son ensemble. »

Le guide proposé par Etalab décline ainsi l'ensemble des dispositions juridiques qui encadrent la prise de décision individuelle automatisée envers des personnes physiques ou morales, partant de la loi relative à l'informatique, aux fichiers et aux libertés de 1978, de la loi pour une République numérique de 2016, du code des relations entre le public et l'administration, et tenant compte du cadre européen défini par le règlement général pour la protection des données (RGPD). Il a ensuite été complété de fiches pratiques pour aider les opératrices de traitement à tenir compte de l'obligation de mention explicite des décisions prises sur le seul fondement d'un traitement automatisé et accompagner les administrations dans l'inventaire des traitements algorithmiques utilisés pour prendre des décisions individuelles.

Toutefois, l'enjeu parcourant ce guide pratique n'est pas que l'opérationnalisation administrative des règles de droit. Il porte sur la nature de la relation entre les citoyennes et l'État et donne une direction au renouvellement de l'idéal démocratique. En effet, l'exigence de motivation des décisions administratives et la possibilité de les contester restreignent les situations d'abus du pouvoir et participent au consentement des administrés. De la même manière, expliquer individuellement les raisons pour lesquelles un algorithme public a produit un résultat et ouvrir la possibilité de débattre publiquement de ses choix de conception peuvent participer de la confiance dans la transformation des services publics et bâtir une appréhension plus démocratique de l'action publique.

Cependant, quatre années après la publication du guide, dans les rapports ordinaires du pu-

blic vis-à-vis de l'administration, un certain nombre de barrières informatiques, juridiques et sociales entravent encore l'explicabilité des algorithmes publics. Nous en avons nous-mêmes fait le constat lorsque nous avons cherché à comprendre le fonctionnement de l'algorithme de calcul des aides au logement [Merigoux, Alauzen et Slimani 2023] que gère la Caisse nationale d'allocations familiales (CNAF).

Or, ces barrières ne relèvent pas nécessairement d'une forme intentionnelle d'opacité de la part des administrations et ne peuvent être levées par la seule coercition (de la juge administratif ou du politique). En effet, la complexité des règles appliquées par les algorithmes publics s'avère d'abord un défi quotidien pour celles et ceux qui les conçoivent et les mettent à jour. Et les exemples sont pléthores : des systèmes de gestion de la paie, au versement des prestations sociales, la combinatoire des règles qui s'enchevêtrent dans les dispositifs administratifs les plus ordinaires s'avère une lourde contrainte. Ainsi, le rapport d'information parlementaire de Gosselin-Fleury et Meslot [2013] sur la faillite du système de paie automatisé, Louvois, conçu par le ministère des Armées, relève parmi les causes de cet échec industriel [Monin 2018], l'insuffisance des spécifications du « calculateur » de paie censé modéliser les règles d'attribution des primes aux soldates. En outre, le rythme auquel ces algorithmes doivent imprimer les changements législatifs et réglementaires est un facteur déstabilisant pour les agents des directions des systèmes d'information. Dans sa réponse à l'interpellation du collectif Changer de cap [2022], la CNAF [2023] se justifie, par exemple, en rappelant qu'il « [leur] est demandé de mettre en œuvre de très nombreuses réformes réglementaires et dans des délais extrêmement brefs qui mettent sous tension [leur] système d'information ».

Outre cette complexité intrinsèque aux montages juridiques et administratifs, les algorithmes publics évoluent suivant des objectifs assignés par les politiques publiques. Or, ces objectifs se transforment dans le temps et leur maintenance devient un défi technique de première importance. Un système dont l'architecture et la mise en œuvre ont été conçues pour une spécification particulière est difficile à faire évoluer dans le temps. Ce phénomène, expliqué par Bellotti [2021] dans son guide à l'usage des informaticiennes dans les directions des systèmes d'information des services publics (et des entreprises), se retrouve, en France, aussi bien dans la gestion du calcul de l'impôt sur le revenu [Merigoux, Monat et Protzenko 2021], qui évolue au gré des lois de finances, que des demandes d'individuation des prestations sociales, dont celle de l'allocation adulte handicapé (AAH) un temps jugée informatiquement impossible [Assemblée Nationale 2021].

À partir d'une revue de la littérature scientifique et institutionnelle et d'une enquête, à la fois informatique et sociologique, sur les formats de l'explicabilité algorithmique, ce rapport de recherche propose d'élargir la conception de l'explicabilité des algorithmes publics, afin de la rendre plus opérationnalisable par les administrations. Nous affirmons que l'explicabilité des algorithmes publics n'est pas seulement un enjeu externe (de redevabilité vis-à-vis des personnes physiques et morales impliquées par ces traitements administratifs, ni même vis-à-vis de la société politique dans son ensemble), mais qu'il s'agit, dans le même temps, d'un enjeu interne à l'administration. La production automatisée de motivations précises et référencées des décisions des systèmes d'information relève d'un enjeu technique indispensable pour maintenir et faire évoluer les algorithmes publics au gré des nouvelles exigences législatives et réglementaires. En d'autres termes, nous défendons l'idée selon laquelle l'explicabilité algorithmique est un point de passage obligé non seulement du contrôle externe de l'administration, mais aussi du contrôle interne de l'administration lui permettant d'être plus transparente à elle-même.

Ce rapport de recherche est structuré en trois parties. D'abord, en dressant un état de l'art, nous esquissons une typologie des formes contemporaines d'explications algorithmiques et de leurs effets sur le public et l'administration. Deuxièmement, nous avons mené une enquête sur l'explicabilité du calcul des aides au logement auprès des actrices de la décision d'attribution et de la contestation de cette décision. Troisièmement, pour abonder le débat, nous avons conçu, dans le langage de programmation Catala [Huttner et Merigoux 2022; Merigoux, Chataing et Protzenko 2021], trois prototypes visant à expliquer les résultats des algorithmes publics dont le comportement est régi par le droit par l'explicitation exhaustive et circonstanciée des règles juridiques s'activant au cours de l'exécution du traitement.

2 Typologie des formes d'explications des algorithmes et des entraves à la redevabilité

Afin de donner à comprendre les obstacles qui entravent la transparence des algorithmes publics, nous avons élaboré une typologie des dispositifs d'explicabilité algorithmique existants dans la littérature scientifique et dans la pratique administrative. Ces dispositifs ont été classés suivant des gradients d'explicabilité pour le public et pour l'administration concernée.

2.1 La transparence du code source et l'explication globale par les règles

Le premier niveau d'explicabilité d'un algorithme public réside dans la transparence de son code source ou des modèles qu'il emploie, ce qui suppose leur communication publique, sur un dépôt d'archive ou un autre support dédié. À cet effet, l'article L300-2 du code des relations entre le public et les administrations précise explicitement que le code source est un document administratif, faisant l'objet d'une obligation de publication en ligne ou de communication sur demande selon l'article L311-1 de ce même code. La communication ou publication des documents administratifs est toutefois soumise à de nombreuses limitations (L311-2, L311-5 et L311-6 du même code), sur lesquelles nous ne nous attarderons pas ici. Deuxièmement, l'article R311-3-1-2 du code des relations entre le public et l'administration dispose une seconde obligation de communication plus précise portant sur les modèles et leurs données.

Le code source Selon le projet GNU, qui fait référence dans les communautés professionnelles du logiciel libre, le code source « se définit comme la forme préférée sur laquelle faire des modifications. Ainsi, la forme qu'un développeur modifie pour développer un programme est le code source de cette version particulière ». Cette définition, fidèle à la pratique, présente l'avantage de faire ressortir immédiatement le lien entre la capacité de comprendre ce que fait un programme et la capacité de lui apporter des modifications. Le code source est donc un vecteur d'explicabilité algorithmique dans le sens où c'est le document auquel la développeuse du programme se réfère afin d'en modifier le comportement. En théorie, une observatrice extérieure peut déduire de la lecture du code source tous les choix de conception de l'algorithme sous-jacent au programme et suivre, pas à pas, les étapes menant à la formation du résultat de l'exécution. En pratique, le code source d'un programme n'est jamais suffisant pour expliquer complètement le résultat de l'exécution de ce programme. Cette exécution dépend également du comportement du langage de programmation, comportement dont la norme peut être fixée de différentes façons, comme le rappelle Grimmelmann

[2023].

Cette subtilité est suffisante pour entraver la redevabilité algorithmique en elle-même. En effet, lors de l'ouverture des codes sources du logiciel CRISTAL de calcul de 24 prestations sociales gérées par la CNAF [Berne 2018] et de calcul de l'impôt sur le revenu [Berne 2016], il est apparu qu'ils étaient écrits dans des langages de programmation inhabituels, respectivement M et COBOL. M est un langage de programmation « propre » à la direction générale des finances publiques (DGFiP) et celle-ci a refusé de publier le compilateur originel de M, mettant en avant des raisons de sécurité, ce qui empêche de reproduire l'exécution du programme. COBOL est un langage de programmation conçu dans les années 1960 et utilisé dans de nombreuses grandes organisations. Cependant, il n'est plus enseigné depuis une vingtaine d'années et rares sont les programmeuses capables de comprendre les programmes écrits dans ce langage. De plus, les environnements de compilation et d'exécution de COBOL sont quasiment tous propriétaires et la profusion de dialectes de ce langage rend impossible la reproduction de l'exécution du programme publié. Dans ces deux cas, la redevabilité algorithmique s'en trouve fortement limitée, car il est impossible de déduire du code source le comportement réel du programme, c'est-à-dire les résultats qu'il calcule à partir d'une entrée donnée. Ainsi, l'exigence légale de communication ou de publication en ligne du code source n'est pas suffisante pour atteindre l'objectif de redevabilité.

Dans d'autres domaines de la production logicielle, le standard de transparence suppose que le code source soit reproductible. Par exemple, dans la recherche en informatique [Association for Computing Machinery 2020], la reproductibilité de l'artefact logiciel est évaluée par les pairs de la même manière que la qualité scientifique d'un article. L'avantage d'un artefact logiciel reproductible est de pouvoir observer les effets des modifications du code source sur les résultats du programme. Ainsi, le code source n'est plus une publication statique, il devient un objet interactif qui permet à une personne de s'approprier le comportement du logiciel et, en quelque sorte, de se mettre à la place de la développeuse. Cette vision de la redevabilité hérite de la philosophie du logiciel libre, et est déclinée au sein du plan d'action pour les logiciels libres et les communs numérique de la DINUM.

Cependant, la stricte reproductibilité technique du calcul à partir du code source d'un programme et de son environnement d'exécution n'est pas non plus suffisante pour constituer une réelle redevabilité algorithmique pour le public ni pour l'administration. En effet, l'interaction avec le code source d'un programme n'est possible que pour les personnes disposant d'une haute compétence en informatique et familières avec les technologies et les langages de programmation utilisés localement. Cette critique qui rejoue la « fracture numérique » est largement étayée par Krajewski [2023]. Elle correspond également à ce que Burrell [2016] nomme « l'opacité par analphabétisme technique », soit le fait que la lecture et l'écriture supposent une compétence spécialisée et soient adressées à des machines.

Outre la question de la compétence nécessaire à la lecture et à l'exécution d'un code source, une critique supplémentaire de la seule publication des codes sources prend, dans la littérature informatique le nom d'obfuscation (ou opacification) du code source. L'obfuscation consiste en une série de pratiques d'écriture logicielle, contraires aux bonnes pratiques de maintenance et de production, qui limitent la compréhension du comportement du logiciel à partir de la lecture du code source, sans toutefois changer le comportement du logiciel au cours de l'exécution. Elle peut être pratiquée à des fins malicieuses, par exemple pour introduire des failles de sécurité difficilement détectables dans un logiciel open source, mais elle peut également être le produit d'une mauvaise ingénierie logicielle, de problèmes de maintenance ou d'une absence de volonté de la développeuse de rendre son code source lisible

pour d'autres. Par exemple, les noms de variables du code source de l'impôt sur le revenu disponible en ligne sont difficiles à déchiffrer, sauf à avoir une connaissance interne du calcul de l'impôt. D'autre part, le code source du calcul des cotisations sociales par l'URSSAF est implémenté dans un langage de requêtes pour bases de données, SQL, avec 10 000 lignes de code faisant usage de fonctionnalités peu usitées d'un langage dont l'utilisation courante ne fait jamais apparaître des requêtes de plus d'une dizaine de lignes de code. Ainsi, l'obfuscation constitue également un défi pour les directions des systèmes d'information qui doivent parfois maintenir des programmes dont les codes sources ne sont plus compris par les agents.

Pour recouvrer la capacité de compréhension du code source, il faut alors consulter ou recréer une documentation externe du code. Cette documentation participe de fait de l'objectif de redevabilité algorithmique et peut être considérée comme étant un document administratif. À ce titre, elle peut être communiquée à toute personne qui en ferait la demande. Bien que des éléments de doctrine juridique comme l'avis CADA n°20181891 à propos de la documentation technique (spécification) du logiciel CRISTAL de la CNAF montrent que cet accès reste, en pratique, à la discrétion de l'administration concernée pouvant faire valoir que « compte tenu de l'ancienneté et de la complexité du système [CRISTAL], l'extraction sollicitée n'était techniquement pas possible sans effort disproportionné ».

Même imparfaite dans la mesure où elle n'est utilisable que par un nombre réduit de personnes (les informaticiennes professionnelles), la transparence du code source des algorithmes publics, la reproductibilité des artefacts logiciels et la mise à disposition de la documentation technique constituent une étape indispensable à la redevabilité algorithmique. A minima, on peut considérer à partir de l'état de l'art en informatique que la transparence du code source de bonne qualité, sans obfuscation, est un objectif désirable pour les administrations elles-mêmes puisque ces techniques participent de la bonne ingénierie logicielle et de la qualité de maintenance des systèmes d'information.

Toutefois, dans la pratique, les administrations opposent des justifications de sécurité aux mesures de transparence ; elles craignent des attaques et des manipulations à partir de qui serait publié en ligne. S'il existe en effet un travail de sécurisation nécessaire et préalable à la publication d'un code source afin d'éviter les attaques informatiques (comme rappelé par le 5° de l'article L311-5 du code des relations entre le public et l'administration), ce travail devrait en toute bonne foi être effectué au fur et à mesure des évolutions du code, car les éventuelles attaquantes n'attendront pas que le code source soit publié pour tenter d'accéder et de compromettre le système.

Les règles définissant les principaux traitements algorithmiques Pour dépasser les contraintes du code source en termes de redevabilité algorithmique, l'article L312-1-3 du code des relations entre le public et l'administration impose une exigence de transparence supplémentaire aux grandes administrations utilisant des programmes informatiques pour produire des décisions individuelles. Comme le rappellent Roy [2023] mais surtout Huttner [2022] avec laquelle nous partageons nombre de conclusions, cette exigence concerne la publication des « règles définissant les principaux traitements algorithmiques », mais uniquement si trois conditions sont remplies. L'administration doit avoir au moins 50 salariées, les traitements algorithmiques doivent être utilisés pour l'accomplissement d'une mission de service public et doivent fonder une décision individuelle. Contrairement au code source qui est un objet informatique bien défini et dont les contours peuvent être décrits avec précisions, la formulation de ce qui doit être publié est extrêmement floue et laisse en réalité beaucoup

de marge à l'administration pour déterminer ce qu'elle entend ou non publier. En effet, la formulation évoque ce qui pourrait se rapprocher de la documentation technique ou de la spécification d'un programme informatique. Cependant, l'adjectif « principaux » semble ici déconseiller en creux aux administrations de se lancer dans un travail de collecte et de publication de la spécification complète de leurs systèmes d'information. Mais alors, que publient en pratique les administrations pour rendre compte des règles définissant les principaux traitements algorithmiques ?

Dans le paysage administratif français, Pôle Emploi fait figure d'exemple en publiant de manière exhaustive sur son site Internet des fiches explicatives, courtes et intelligibles, relatives aux traitements algorithmiques que l'agence opère sur les situations des millions de demandeuses d'emploi. La lecture de ces fiches fournit un premier niveau d'explication générale sur le traitement et ses grandes étapes et rend les vérifications par les bénéficiaires possibles. Dans le cas du calcul du montant et de la durée de l'allocation de retour à l'emploi (ARE), la fiche de Pôle Emploi est suffisamment complète pour permettre à une demandeuse d'emploi de refaire le calcul de ses droits — ou, plus probablement, pour qu'une médiatrice (assistante de service social, membre d'une association, avocate, etc.) procède à une vérification de la prise en compte des paramètres de sa situation.

Cependant, ce dispositif souffre, y compris dans sa version la plus aboutie telle qu'appliquée par Pôle Emploi, du décalage entre les « règles définissant les principaux traitements » et la spécification complète de ces traitements. La complexité du droit pris en compte dans le calcul automatique, ici celui du calcul des indemnités chômage, apparaît lorsque l'on s'écarte de la situation idéale typique (une salariée précédemment embauchée en CDI) et que l'on s'aventure dans les carrières régies par un droit spécial et procédant de nombreuses exceptions (étrangères, régimes spéciaux, etc.). Par exemple, dans les cas complexes avec des droits particuliers, il est possible que l'administration n'ait même pas défini de règle à suivre, ou que ces dernières ne soient pas connues ou interprétées de manière univoque par les agents, voire qu'elles fassent l'objet de contestations [Gaudin 2019]. La publication d'une spécification complète aurait alors pour mérite de clarifier la position de l'administration dans tous les cas et, ce faisant, bénéficierait y compris à ses propres agents.

La fiche explicative des règles définissant les principaux traitements apparaît alors relever davantage de la communication institutionnelle et d'une démarche pédagogique vis-à-vis du public que d'un document juridiquement opposable, renforçant les droits personnels et clarifiant les processus internes. Le caractère court et incomplet de ce type de fiches se révèle alors à double tranchant : première source d'information intelligible sur le traitement pour la demandeuse d'emploi, la fiche ne permet toutefois pas de faire valoir ses droits en toutes situations ni d'améliorer le fonctionnement des services publics. C'est ainsi que ce genre de dispositifs peut s'apparenter à la logique de démonstration, telle que mise en lumière par Rosental [2019] et illustrée par Cath et Jansen [2021] : un cheminement écrit ou audiovisuel dont la vocation affichée est d'ordre probatoire ou pédagogique, mais qui remplit ce faisant nombre d'autres rôles, comme se donner en spectacle, tout en cachant des secrets de fabrication, gérer des projets de transformation d'organisations, démarrer une transaction, créer du lien social, etc. Dans le cas des projets de transparence des administrations, de la LOLF hier aux récents hackathons d'ouverture des données et des codes sources publics, les sceptiques peuvent être enclins à penser qu'il s'agit d'une mise en spectacle de la transparence de l'État, plutôt que la refondation du lien de confiance avec les citoyennes.

Ainsi, l'état des lieux de ce premier niveau d'explicabilité algorithmique de la transparence des codes sources et des règles définissant les principaux traitements est ambivalent. Des

exigences légales ont consacré un droit d'accès limité, mais réel ; or, celui-ci est entravé par des obstacles techniques, juridiques et politiques. Et, même s'ils étaient tous levés, les codes constitutifs des algorithmes publics resteraient inaccessibles à la plupart des personnes concernées par ces décisions algorithmiques, dans la mesure où ces dernières n'ont pas, par défaut, de compétences pour lire et interpréter ces informations techniques. En effet, le phénomène a déjà été décrit par Goëta et Davies [2016] à propos de l'ouverture des données publiques : ces informations supposent une lecture par des machines plus que par des citoyennes ou des usagères ordinaires du service public. En ce sens, l'ouverture ne peut être qu'un fondement technique, autrement dit une condition nécessaire mais non suffisante de la redevabilité algorithmique (ou, dans les termes de Huttner [2022], une fonction technique). La publication des règles définissant les principaux traitements opère d'ailleurs ici un compromis qui vient tenter d'améliorer l'intelligibilité du code source, mais l'utilité de cette publication complémentaire reste à prouver par l'usage.

2.2 Les formes d'explication du calcul par l'administration

Face à l'inintelligibilité du code source et le manque de détails des documents pédagogiques décrivant le traitement, la législatrice a prévu un second dispositif technique d'explicabilité des algorithmes : le droit pour les personnes concernées à une double explication, générale et individualisée, du traitement algorithmique, qui a été mis en œuvre à leur égard. Il s'agit là d'un événement dans l'histoire administrative dans la mesure où, au cours des 30 dernières années, un écart s'est creusé entre, d'un côté, l'automatisation des décisions individuelles par le déploiement de systèmes d'information dans tous les services publics et, de l'autre, la production de justifications de ces décisions, qui a continué à s'opérer souvent a posteriori et au cas par cas. En conséquence, le droit à l'explication détaillée et intelligible ouvre la possibilité de réaligner le traitement d'une décision individuelle et son explication ajustée à la personne concernée.

En général : les règles et principales caractéristiques du traitement L'article L311-3-1 du code des relations entre le public et l'administration établit deux dispositifs essentiels pour la redevabilité algorithmique. Le premier est purement informatif et d'ordre général. Il s'agit de la notification à l'usagère dont la situation a été traitée par un algorithme que celle-ci a été exposée (totalement ou partiellement) à une prise de décision automatisée. Cette notification porte à la connaissance de particuliers le processus de décision mis en œuvre par une administration donnée. L'exigence paraît simple à satisfaire, mais sa mise en œuvre présente un premier défi à l'administration puisqu'elle demande de réaliser un inventaire de tous les algorithmes utilisés par les services de l'administration. Il s'agit donc de mener un important travail de communication et coordination interne à l'administration pour enrôler tous les services opérateurs d'algorithmes prenant des décisions automatisées dans une démarche de notification à leurs utilisatrices. Datactivist [2022] propose une méthodologie pour réaliser cette démarche, qui a été suivie à la Métropole européenne de Lille [Donzel et al. 2022] : « elle contribue à solidifier voire à améliorer la procédure décisionnelle car elle oblige les services à s'interroger sur leurs méthodes et leur organisation. Elle permet également d'acculturer à l'usage et à la redevabilité d'un algorithme public ». La même ligne de travaux a mené à la parution d'un cahier sur la transparence des algorithmes publics écrit par Banuls et al. [2023] dont l'autrice principale participe également au présent rapport.

Ensuite, l'article exige que l'administration communique, en cas de demande de l'usagère, un document explicatif intelligible des « règles définissant ce traitement, ainsi que les prin-

cipales caractéristiques de sa mise en œuvre ». À l'instar des « règles définissant les principaux traitements algorithmiques », cette formulation ne fait mention d'aucun objet informatique tangible, de sorte que sa formulation n'oblige pas l'administration à coucher par écrit la spécification complète du traitement algorithmique. Cependant, l'article R311-3-1-2 vient préciser cette formulation et la décliner sous quatre items distincts.

Premièrement, le « degré et le mode de contribution du traitement algorithmique à la prise de décision » indique si la décision a été prise de manière entièrement automatisée ou non, ce qui est essentiel, mais ne décrit pas le détail de l'éventuel calcul sur lequel la décision est fondée.

Deuxièmement, les « données traitées et leurs sources » constituent un élément basique et indispensable de la redevabilité algorithmique : elles permettent de connaître la manière dont le système informatique a été alimenté (de nombreuses erreurs proviennent de la saisie de données erronées).

Troisièmement, « les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé » se confondent en partie avec les données traitées, auxquelles il faut rajouter des variables intermédiaires créées par le calcul. Le mot « pondération » (poids dans un modèle d'apprentissage supervisé) se réfère plus spécifiquement aux techniques d'apprentissage statistique, même si les modèles actuels sont beaucoup plus complexes qu'une simple pondération affectée à chacun des paramètres.

Enfin, les « les opérations effectuées par le traitement » sont, là encore, une formulation floue ne correspondant pas exactement à une spécification ni à une documentation technique — même si l'obligation n'est pas ici suivie de l'adjectif « principal », ce qui peut laisser penser qu'on est en droit d'exiger du détail.

Comme évoqué ci-dessus, certains algorithmes publics se distinguent par l'utilisation de techniques d'apprentissage statistique dans leurs programmes (plus couramment appelées intelligence artificielle, sur cette qualification voir Cardon, Cointet et Mazières [2018]). De nombreux travaux ont mis en évidence que ces techniques d'apprentissage statistique se caractérisent par un genre inédit d'opacité : Burrell [2016] soutient notamment la thèse selon laquelle leur opacité découle de l'inadéquation entre l'optimisation mathématique en haute dimensionnalité, caractéristique de l'apprentissage automatique, et les exigences d'un raisonnement compréhensible par des humains et des styles d'interprétation sémantique. Paradoxalement, même s'ils sont moins répandus dans l'administration et utilisés de manière plus récente que les programmes à base de règles logiques, l'exigence légale de transparence de leur code source est bien plus précise et exigeante. En effet, l'article R311-3-1-2 du code des relations entre le public et l'administration intègre la particularité de ces techniques d'apprentissage statistique pour lesquelles le code source est insuffisant pour rendre compte du comportement du programme. Aussi, il exige la publication des données utilisées pour l'entraînement des modèles statistiques, ainsi que la liste des paramètres formant l'espace dans lequel ces données et leur pondération évoluent.

Cette disposition législative peut être lue comme étant une déclinaison de l'objectif de redevabilité au cas particulièrement scruté de ces algorithmes basés sur l'apprentissage statistique. Toutefois, ces dispositions peinent à s'établir et les pratiques de publication demeurent variables. Par exemple, le projet-phare de start-up d'État « Signaux Faibles », utilisant un modèle d'apprentissage statistique de détection des entreprises en difficulté à partir de leurs données déclarées à la DGFiP et à l'URSSAF, respectait bien l'article R311-3-1-2 avec un code source et une documentation exhaustive de son modèle. Mais ces derniers ne

valent que jusqu'à la fin de l'année 2021. En effet, à son intégration aux services de la DGFiP, les agents n'ont pas poursuivi l'effort de publication initié par la start-up d'État. D'autre part, le modèle statistique de détection des fraudeurs utilisé par les CAF a fait l'objet d'une demande de documents administratifs par La Quadrature du Net [2023b], or, pour l'instant, le code source transmis pour le modèle actuel fait l'objet d'une obfuscation volontaire, qui escamote la liste des paramètres du modèle. La Quadrature du Net [2023a] a néanmoins pu récupérer et analyser le code source des modèles utilisés entre 2010 et 2018. Le manque de transparence des modèles d'apprentissage statistiques et des données utilisés par l'administration n'est pas un problème spécifiquement français. Aux Pays-Bas, la question a été l'objet d'un débat national après un scandale lié à une liste noire secrète de ménages suspectés de fraude (souvent à tort) qui circulait entre les administrations [Geiger 2021], et la presse internationale se fait l'écho de cas comparables dans la plupart des pays disposant de bureaucraties modernes.

Individuellement : une explication détaillée et intelligible Un dispositif d'explication plus exigeant que les principales caractéristiques du traitement est requis dans le 2° de l'article 47 de la loi informatique et libertés : le responsable du traitement ou l'administration dans les cas qui nous intéressent, doit « pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard ». Cette formulation ouvre la voie à un droit à la redevabilité algorithmique ambitieux supposant la combinaison de trois éléments :

1. une explication individualisée, requérant l'exposé de l'exécution du programme sur le cas d'espèce, plutôt que du fonctionnement général du programme (comme dans le cas du code source de la section 2.1) ;
2. une explication détaillée, soit une exigence de complétude de l'explication, supposant de justifier chaque étape de traitement effectuée par le programme, ce qui pousse l'opérateur du traitement à un supplément de transparence par rapport aux principales caractéristiques du traitement de l'article L311-3-1 du code des relations entre le public et l'administration ;
3. une explication intelligible, contenant tous les éléments permettant à la personne récipiendaire de la comprendre.

Ces trois caractéristiques sont les piliers d'un mode d'explication dont nous n'avons pas trouvé d'illustration dans les pratiques administratives [1], mais qui pourrait, à l'avenir, s'avérer bénéfique à la fois aux personnes concernées par les traitements automatisés et par les administrations qui deviendraient plus transparentes à elles-mêmes. Ainsi, l'alinéa 2 de l'article 47 de la loi informatique et libertés étend la logique de l'article L211-5 du code des relations entre le public et l'administration. Cet article exigeant la motivation juridique des décisions administratives dans le cas où la décision est défavorable (cas listés par l'article L211-2 du code des relations entre le public et l'administration) est l'un des fondements de la redevabilité administrative et de l'État de droit : la motivation tient à distance l'arbitraire, elle peut être contestée et débattue, y compris devant le tribunal administratif. Ce principe général est réitéré et précisé dans le cas du versement des prestations sociales aux articles L211-7 et L211-8 du code des relations entre le public et l'administration (sur ce point, voir Huttner [2022]).

1. Ni dans la doctrine des institutions juridiques ; par exemple, l'importante étude du Conseil d'État [2022] sur l'intelligence artificielle cite bien page 348 l'alinéa 2 de l'article 47 de la loi informatique et libertés mais ne la relie pas aux obligations d'explicabilité des systèmes d'intelligence artificielle publics, pourtant explicitées page 122 et suivantes.

Une telle motivation des décisions administratives est institutionnalisée ; elle dispose de ressources humaines et matérielles spécifiques (directions des affaires juridiques, services de médiation, etc.). Elle est aussi une compétence distinctive des agents publics formés aux différentes branches droit administratif. Or, la production de ces motivations se fonde sur un traitement méticuleux et au cas par cas des dossiers, qui transitent ensuite le long d'une chaîne de lecture et d'écriture engageant, par la signature de la décision, l'ensemble de l'administration concernée. Aussi, même quand on dispose d'un inventaire de cas types et de modèles de motivations associées, la sociologie du travail [Denis et Pontille 2012] a bien montré que l'épaisseur intellectuelle, le temps nécessaire, l'invisibilité du travail administratif était toujours plus importante que l'on pourrait le croire. De sorte que la production des motivations des décisions administratives n'a, en pratique, pas de caractère massif (elle ne s'applique pas à tous les dossiers traités automatiquement) et se voit donc reconduite aux barrières d'accès aux droits personnels : accès à une aide juridique, complexité des procédures, etc.

Avant les transformations des services publics visant à décentrer la relation administrative du guichet, le guichet de proximité constituait l'espace d'explication individuelle par excellence (et, le cas échéant, de réparation) de la décision administrative. L'agent au guichet ou lors de rendez-vous était souvent, à la fois, le vecteur par lequel la décision était prononcée, mais aussi celui par lequel l'explication était fournie, et ce, pas uniquement en des termes juridiques. Toutefois, comme le condense Weller [2003], « Composé de ressources matérielles, humaines et symboliques, le guichet apparaît comme un agencement hybride destiné à dire le droit ». Après l'interaction de guichet, si l'usagère n'était pas satisfaite ou constatait un quelconque manquement, elle pouvait ensuite ouvrir une procédure de réclamation, voire attaquer la décision de l'administration.

Au fil de la réduction de l'accueil de proximité des usagères des services publics et de la multiplication des téléprocédures [Deville 2023 ; Lequesne-Roth, Kimri et Legros 2021], cette forme ordinaire d'explication de la décision n'est devenue disponible que pour certains publics ciblés comme étant très fragiles (ou bénéficiant, par exemple, de l'accompagnement dans le temps d'une assistante sociale). Or, cela pose la question de savoir à qui il incombe de réaliser, en masse, le droit des usagères à la motivation des décisions administratives afin que ce dernier ne reste pas lettre morte.

Partant du principe que ces décisions nécessitant des explications en masse sont d'ores et déjà produites par des programmes informatiques mettant en œuvre des algorithmes publics, l'esprit de l'alinéa 2 de l'article 47 de la loi informatique et libertés est d'inciter l'administration à se réformer en concevant des instruments de motivation systématique de ses décisions administratives. Cela implique un renversement du raisonnement : au lieu de penser la motivation comme une opération intervenant par exception et a posteriori dans le cadre d'un contentieux ou d'une réclamation, il s'agit de concevoir la motivation par défaut et en parallèle de la prise de décision administrative. Autrement dit, de tenir ensemble, techniquement, la décision et son explication individualisée, détaillée et intelligible, au sens de l'article 47 de la loi informatique et libertés.

La génération automatisée d'explications de décisions administratives parallèles aux décisions apparaît également comme une promesse d'égalité de traitement. Elle semble porteuse d'un idéal de production par l'État d'une citoyenne apte à appréhender la technicité du droit et, le cas échéant, à le contester. Toutefois, des incertitudes techniques s'opposent encore à l'accomplissement de ce type d'instrument administratif.

Les fictions explicatives des algorithmes publics Pour analyser les obstacles techniques à la génération automatique d'explications individualisées, détaillées et intelligibles des décisions, nous avons passé en revue des formes contemporaines de la motivation administrative. Deux objets administratifs ordinaires s'approchent de l'explication automatique individualisée, détaillée et intelligible de la décision : la fiche de paie et l'avis d'imposition. La décision de l'administration correspond pour la fiche de paie au montant net à payer à la salariée après le prélèvement des cotisations sociales et des prélèvements obligatoires, et, pour l'avis d'imposition, au montant d'impôt dû par le contribuable au titre de l'impôt sur le revenu, ainsi que les contributions assises sur la base de la déclaration des revenus annuelle.

Rubrique de paie libellé	Nombre ou base	Taux	Gains	Retenues	Charges patronales	
					Taux	Montant
011N TRAIT.BASE MENS. NT			2 085,00			
748N Remb. transport			14,25			
844C CRDS	2 048,51	0,500		10,24		
859C CSG déductible	2 048,51	6,800		139,30		
869C CSG non déductible	2 048,51	2,400		49,16		
1811C URSSAF maladie RG	2 085,00				6,000	125,10
811C URSSAF maladie RG	2 085,00				7,000	145,95
479C URSSAF FNAL RG	2 085,00				0,500	10,43
804C URSSAF vieil. pla RG	2 085,00	6,900		143,87	8,550	178,27
806C URSSAF vieil. dép RG	2 085,00	0,400		8,34	1,900	39,62
847C URSSAF autonomie RG	2 085,00				0,300	6,26
1826C URSSAF AF RG TX1	2 085,00				3,450	71,93
1827C URSSAF AF RG TX2	2 085,00				1,800	37,53
827C URSSAF acc. trav. RG	2 085,00				0,700	14,60
840C Retraite Ircantec A	2 085,00	2,800		58,38	4,200	87,57
848C AOT taxe transp. RG	2 085,00				2,950	61,51
865C Pôle empl. chômage A	2 085,00				4,050	84,44
829C Taxe sur salaires FT	2 085,00				4,250	88,61
849C Taxe sur salaires F2	2 085,00				9,350	194,95
879C Abattement Taxe Sal.	- 283,56				6,000	- 17,01

Figure 1 – Extrait d'une fiche de paie d'un des auteurs du rapport

La fiche de paie, illustrée par la Figure 1, vise à expliquer la soustraction opérée sur le montant du salaire brut négocié entre l'employeuse et la salariée, ayant abouti au salaire dit net à payer, qui est versé sur le compte en banque de la salariée pour un mois donné. La présentation d'une telle fiche de paie constitue une obligation légale (article L342-3 du code du travail) de l'employeuse dont le contenu est précisément réglementé (article R3243-1 du code du travail). En effet, la fiche de paie telle qu'elle se donne à lire dans sur une page A4 est le produit d'une myriade de décisions administratives correspondant à chacun des prélèvements salariaux ou patronaux effectués sur le salaire brut. Ces décisions prises tous les mois pour le compte de millions de salariées ont une incidence financière immédiate, dans la mesure où elles affectent le salaire perçu. Lorsqu'une salariée constate un montant de salaire encaissé trop faible ou trop élevé par rapport à son estimation, elle peut se reporter à la fiche de paie (ou la faire lire par une déléguée syndicale ou une personne compétente) afin d'obtenir une explication individualisée, détaillée et intelligible des prélèvements qui lui ont été imposés. Sans cette médiation, la salariée devrait interroger son employeuse ou la personne responsable des ressources humaines pour obtenir une telle explication. Grâce à ce document formalisé et délivré mensuellement, la salariée peut lire (ou faire lire) ce qui est prélevé sur son salaire brut et ne contacter l'employeuse qu'en cas de réclamation. Elle dispose alors de prises textuelles pour argumenter la réclamation.

Si la fiche de paie constitue un objet administratif portant en même temps et automatiquement une décision administrative et son explication, l'intelligibilité de cette explication reste perfectible : malgré des modifications régulières justifiées, par exemple, par une quête de lisibilité parfois confiée à des designers [Gélédan 2021], beaucoup de salariées se révèlent incapables de détecter les erreurs dans leur fiche de paie [Gérard 2021]. Cela résulte en partie de la relation de confiance établie avec l'employeuse, qui se traduit par défaut par une délégation de la vérification aux professionnelles de la comptabilité et des ressources humaines. Mais, on peut aussi suivre l'hypothèse des designers qui se sont penchées sur le format d'explication graphique et postulent une limitation de l'intelligibilité. L'explication se présente en effet sous la forme d'une série compacte de lignes. Chaque ligne contient l'assiette d'un prélèvement, son taux et la somme prélevée. Aucune justification textuelle ou symbolique de chaque ligne ne figure sur la fiche de paie, ne permettant de comprendre sans une lecture professionnelle ni le taux ni l'assiette considérée.

Détail des revenus	**Déclar. 1**			**Total**
Salaires	24252			
Déduction 10% ou frais réels	- 2425			
Salaires, pensions, rentes nets	21827			21827
Revenu brut global				**21827**
Revenu imposable				**21827**
Impôt sur les revenus soumis au barème [14]				1276
Décote				- 213
Impôt total avant crédits d'impôt			**1063**	
CREDITS D'IMPOT, IMPUTATIONS	**Montant déclaré**		**Montant retenu**	
Cotisations syndicales	30		30	
Montant du crédit d'impôt calculé				- 20
IMPOT NET				
Total de l'impôt sur le revenu net				1043
CALCUL DU SOLDE DE VOTRE IMPOT POUR 2021 :				
IMPOT SUR LE REVENU				
Impôt sur le revenu 2021 dû [53]:				1043
Retenue à la source prélevée en 2021 par vos verseurs de revenus :				- 1659
Solde d'impôt sur les revenus 2021 :				**- 616**
COMPTE TENU DES ELEMENTS QUE VOUS AVEZ DECLARES, LE MONTANT QUI VOUS SERA REMBOURSE (voir notice) EST DE				616
CE REMBOURSEMENT EST AUTOMATIQUE, VOUS N'AVEZ AUCUNE DEMARCHE A FAIRE.				
INFORMATIONS COMPLEMENTAIRES				
Revenu fiscal de référence [25]				**21827**

Figure 2 – Extrait d'un avis d'imposition d'un des auteurs du rapport

L'autre objet qui nous paraît s'approcher le plus d'une production automatique d'explications individualisées, détaillées et intelligibles parallèle à la décision administrative est l'avis

d'imposition émis par la Direction générale des finances publiques (DGFiP) au titre de l'impôt sur le revenu. La figure 2 est une illustration du format de l'avis d'imposition de l'année 2022, à partir du cas le plus simple : une seule source de revenus salariale pour un foyer fiscal composé d'un seul individu. L'avis d'imposition est un document réglementé notamment par l'article 170 du code général des impôts et l'article L253 du livre des procédures fiscales. La DGFiP a choisi d'aller beaucoup plus loin dans l'explication du détail des calculs que ce qui est strictement requis par le droit. Ainsi, la production de l'avis d'imposition (auquel la DGFiP joint un mode d'emploi, des vidéos pédagogiques et une infographie sur les recettes de l'État et la dépense publique) est automatisée et tire ses informations du moteur de règles fiscales décrit par Merigoux, Monat et Protzenko [2021]. Ce système de la DGFiP est remarquable dans le sens où il permet de restituer n'importe quelle variable intermédiaire du calcul, dont la valeur est affichée (si elle est présente) dans l'avis d'imposition. En ce sens, l'avis d'imposition présente une explication beaucoup plus détaillée que celle de la fiche de paie, même si elle ne représente pas l'intégralité du calcul. En effet, le calcul du montant d'impôt à partir de la déclaration des revenus nécessite de recalculer plusieurs fois le montant de l'impôt à partir de déclarations fictives : par exemple, pour calculer le plafonnement des avantages du quotient familial, il faut calculer l'impôt avec et sans les personnes à charge du foyer fiscal. C'est ce que l'on appelle une double liquidation, et qui se retrouve dans d'autres algorithmes publics comme celui consacré au calcul des aides au logement [Merigoux 2022]. Or, dans son avis d'imposition, la DGFiP produit les valeurs des variables intermédiaires uniquement pour la dernière liquidation, censée contenir tous les bons résultats choisis grâce aux liquidations précédentes. En ce sens, l'avis d'imposition est une fiction explicative, visant à représenter de manière arithmétiquement exacte un calcul plus complexe du point de vue du programme informatique.

Si la fiction explicative de l'avis d'imposition peut aller à l'encontre du critère de caractère détaillé de l'explication, elle s'avère être un compromis pratique avec l'intelligibilité de la décision administrative, en l'espèce le calcul de l'impôt sur le revenu des particuliers. Face à un calcul démesurément complexe, reflétant la complexité propre au droit fiscal qui spécifie ce calcul, la création d'une telle fiction retraçant les grandes étapes du calcul de l'impôt est un outil heuristique à destination des contribuables (et des professionnelles concernées), permettant de satisfaire partiellement l'objectif de redevabilité. En effet, cette fiction ne saurait tenir lieu d'explication complète dans la mesure, où d'un point de vue algorithmique, elle manque de détails. On comprend donc à partir de ce cas que la conception d'un format explicatif répondant à la triple exigence d'individualisation, de détail et d'intelligibilité requiert un assemblage *ad hoc* de nature à guider la personne concernée vers une compréhension de techniquement et lisible de la décision qui lui est administrée.

2.3 Des dispositifs d'explication automatiques plus complets et individualisés

Le problème de la production automatisée d'explications individualisées, détaillées et intelligibles de décisions administratives est maintenant posé. Satisfaire pleinement aux exigences de l'alinéa 2 de l'article 47 de la loi informatique et libertés s'avère difficile ; dans la mesure où c'est une tâche qui nécessite de clarifier ce que l'on entend par l'explication, en complément du travail effectué sur le calcul informatique lui-même. Dans cette section, nous esquissons des formes que pourrait prendre l'explication automatique, à l'aide d'une sélection des cas administratifs prototypiques objets de débats dans la littérature scientifique.

Un idéal graphique d'information contextualisée et différenciée En 2018, l'Association pour la Fondation d'un Internet Nouvelle Génération (FING) a mis ligne dans un billet de blog [Guillaud 2018] une série de travaux exploratoires consacrée à la transparence algorithmique, soutenue par Etalab, impliquant des étudiants en design de l'école Boulle et de l'ENS Cachan, encadrés, entre autres, par le spécialiste Cellard [2020]. Ces travaux visaient à concevoir des interfaces de visualisation d'objets complexes produits par les administrations, en revenant au contexte du message d'explication du calcul d'un algorithme et aux dispositifs par lesquels cette explication se trouve présentée à la personne concernée. Parmi les résultats de ces explorations graphiques élaborées au moyen d'outils de visualisation de données, deux formes d'explication concernant le calcul de la taxe d'habitation ont retenu notre attention dans la mesure où leur formalisation nous semblait susceptible de satisfaire le triple objectif légal d'individualisation, de détail et d'intelligibilité.

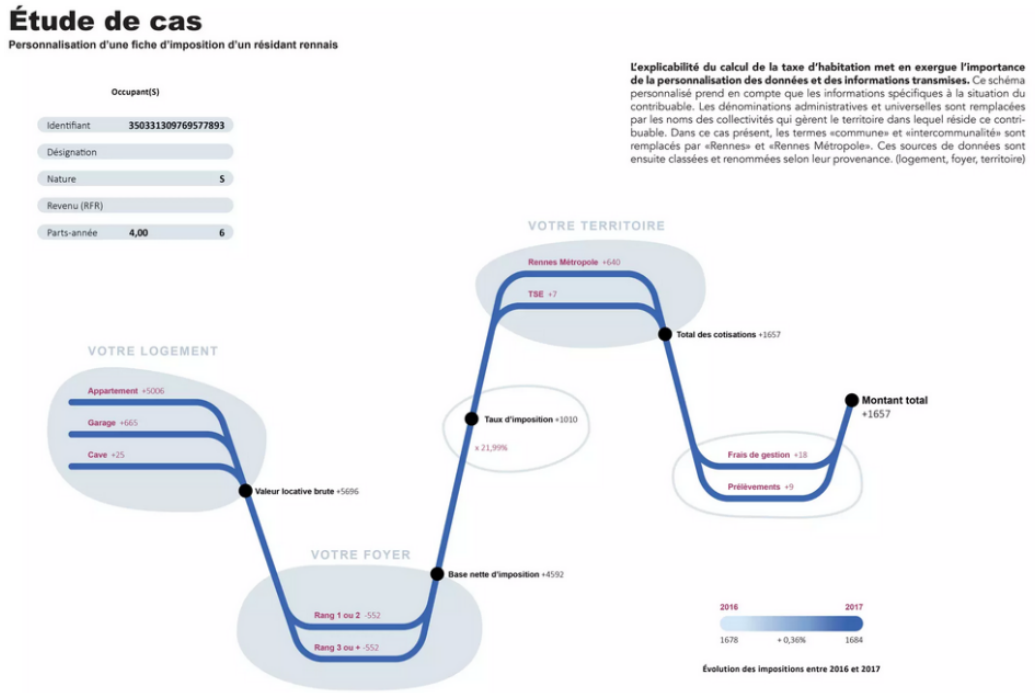


Figure 3 – Visualisation extraite de [Guillaud 2018]

Ces prototypes d'interfaces, illustrés en Figures 3 et 4, sont graphiquement très élaborés. Ils se démarquent des objets administratifs ordinaires comme la fiche de paie ou l'avis d'imposition, qui, par contraste, s'apparentent à un banal tableau comptable dans lequel chaque ligne apporte une information supplémentaire. Mais, cette présentation visuellement plus agréable pour les lecteurs disposant d'une capacité de lecture graphique avancée suppose un coût de conception et de maintenance informatique plus grand, dès lors que l'on considère que l'explication doit être ajustée pour prendre en compte les évolutions du calcul de la taxe d'habitation dont il est question dans l'explication. Les modifications des paramètres de calcul des taxes sont en effet récurrentes dans les lois de finances et les administrations

doivent prendre en compte, tous les ans ou plus fréquemment, les changements législatifs, réglementaires ou administratifs en faisant évoluer les algorithmes qui opèrent le calcul des taxes. Sachant que les administrations ont du mal à maintenir la correction basique de ces algorithmes, qui deviennent au fil du temps d'une extrême complexité (dans le cas des États-Unis, voir Escher et Banovic [2020] et Kennan et Soka [2022]), il apparaît peu probable que ces formes exploratoires puissent donner lieu à un développement informatique visant à maintenir dans le temps un tel dispositif de production d'explications. En cela, bien que produits sous la houlette d'administrations (Etalab), ils se placent dans la lignée de recherches théoriques sur la notion d'explication comme celles de Lombrozo [2006] sur les biais cognitifs liés à l'explication (voir Bertrand et al. [2022] pour une synthèse récente du sujet) ou celles de Kroll et al. [2017] sur le design de systèmes informatiques redevables. Aussi, nous considérerons ces prototypes comme des idéaux heuristiques pour sensibiliser et orienter des actions en faveur de la redevabilité algorithmique, mais pas comme des solutions technologiques durables susceptibles de répondre aux critères légaux de l'explication.

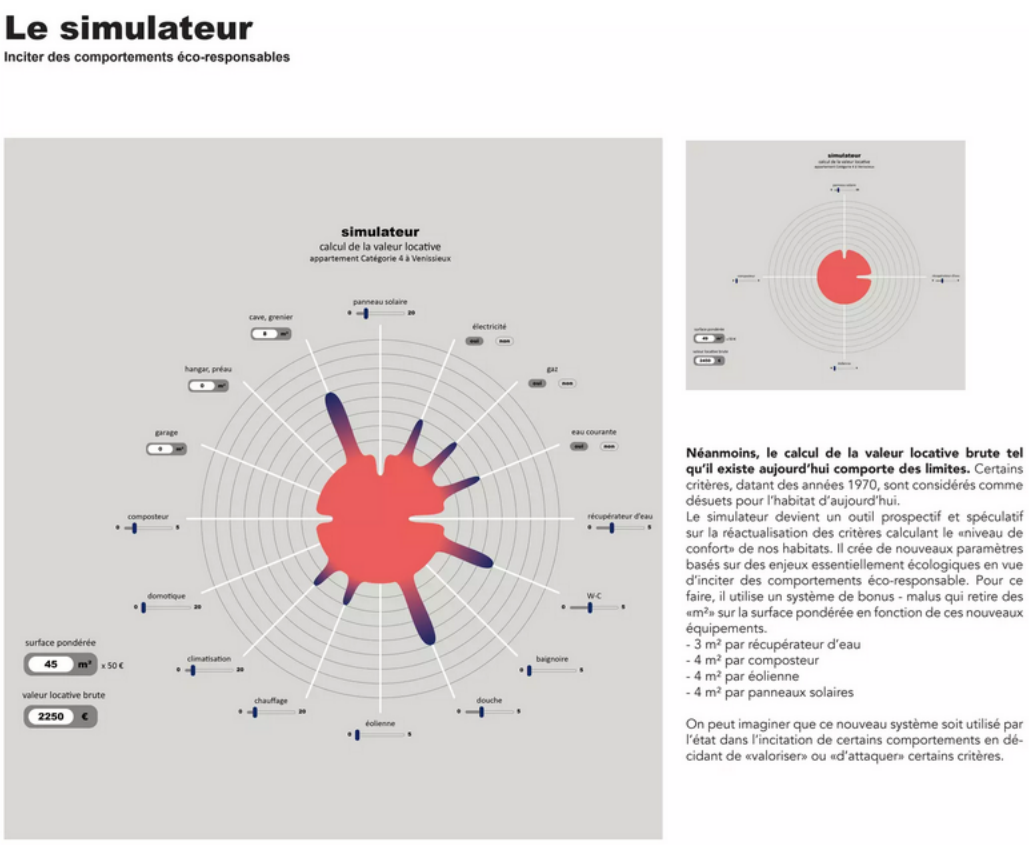


Figure 4 – Visualisation extraite de [Guillaud 2018]

Le défi de l'explicabilité des boîtes noires L'autre travail significatif portant sur un versant plus technique de l'explicabilité algorithmique est celui de Henin [2021], décliné dans

plusieurs publications [Henin et Le Métayer 2021, 2022]. Hénin se focalise sur les scénarios dans lesquels l'algorithme est constitué en boîte noire (*black box*), autrement dit, son code source est indisponible ou inintelligible. Il examine alors les moyens d'extraire du système par de simples requêtes les informations nécessaires à la fabrication d'une explication sur son fonctionnement. Ce scénario de boîte noire correspond principalement aux programmes informatiques s'appuyant sur les techniques d'apprentissage statistiques [Cardon, Cointet et Mazières 2018] dont nous avons évoqué les enjeux de transparence des modèles plus haut 2.2. De manière significative, le choix du scénario en boîte noire, par opposition à celui de la « boîte blanche », est justifié par Vallet et Henin [2022] en évoquant les entraves à la redevabilité algorithmique des codes sources (dont le temps des agents), discutées en section 2.1 :

> « Documents, codes sources, données, etc., les contrôleurs de la CNIL sont ainsi souvent amenés à prendre copie de diverses pièces dans le cadre de leurs investigations. De ce fait, des méthodes d'audit en boîte « blanche » c'est-à-dire se fondant sur l'analyse du code du système et/ou ses paramètres (coefficients du modèle) et des données sont donc envisageables pour les algorithmes d'IA qu'aurait à analyser la CNIL. Cependant, si ces méthodes permettent d'avoir une vue complète et exhaustive du système contrôlé, elles sont par nature complexes à mettre en œuvre. En effet, elles sont bien souvent spécifiques à un système d'IA donné et requièrent d'être en mesure d'exécuter (« faire tourner ») celui-ci, ce qui s'avère bien souvent être une tâche délicate. Configuration à reproduire à l'identique, besoin de puissance suffisante, nécessité d'avoir accès à de nombreuses infrastructures techniques et d'être en mesure de les exploiter, etc. les obstacles sont nombreux. Enfin, les méthodes d'audit en boîte blanche demandent un temps conséquent pour leur mise en œuvre. Aujourd'hui, les agents de la CNIL disposent d'un temps limité pour l'analyse des éléments recueillis en contrôle et l'établissement de constats (généralement inférieur à 3 jours). »

Hénin propose deux outils pour analyser un algorithme en boîte noire. Le premier, IBEX, prend pour point de départ un ensemble de données de couples (entrée, sortie) de l'algorithme et en réalise une analyse selon trois modes. Premièrement, il établit un modèle statistique de l'algorithme selon le jeu de données qui lui permet d'afficher l'importance relative de chaque variable d'entrée sur la sortie. Deuxièmement, à partir d'une décision donnée de l'algorithme, il élabore et teste empiriquement sur le jeu de données des règles de décision symboliques qui pourraient expliquer sous une forme plus générale la décision de départ. Enfin, s'appuyant sur les travaux de Wachter, Mittelstadt et Russell [2017], l'outil génère pour une décision donnée un contre-factuel, c'est-à-dire une autre décision de résultat opposé, mais dont l'entrée est proche de celle de la décision originale. Ainsi, IBEX est à proprement parler un dispositif d'explicabilité qui tente de fournir à son utilisatrice une explication individualisée, détaillée et intelligible. C'est ce genre de dispositifs d'explicabilité qui est promu par Olsen, Slosser et T. T. Hildebrandt [2020]. Dans ce papier, Olsen et al. marquent leur désaccord avec la volonté de fournir automatiquement une explication *causale*, c'est-à-dire inspirée de la preuve mathématique, à des décisions prises automatiquement : la décision impliquerait toujours nécessairement des éléments de qualification subjectifs qu'on ne peut pas expliquer causalement. Il suffirait alors pour un système automatisé de fournir une explication aussi bonne qu'une motivation juridique écrite par un fonctionnaire humain ; le critère d'évaluation du dispositif d'explicabilité automatique étant alors l'indistinguabilité par rapport aux explications rédigées par un humain. Or, nous considérons que cette solu-

tion en scénario boîte noire est excessivement focalisée, d'une part, sur le défi contemporain des algorithmes d'apprentissage automatique et, d'autre part, sur l'arène judiciaire et méconnaît les décisions automatisées massives prises depuis plusieurs décennies déjà, de manière déterministe et connue sur la base seule du droit (une fois le formulaire rempli par l'usagère et traité par l'administration), et pour lesquelles une explication *causale* semble être nécessaire pour fournir une explication détaillée, individualisée et intelligible.

IEEE Std 7000-2021
IEEE Standard Model Process for Addressing Ethical Concerns during System Design

Table B.1—Principles for value ranking

Values are higher…	Examples
… the more they endure (has nothing to do with absolute time, but with the *persistence* of a value, the eternality of a value	Love is higher than enthusiasm; happiness is higher than convenience.
… the less they are extensible or divisible	A piece of art cannot be divided, which is why it is of higher value than a piece of bread; beauty as a phenomenon is of higher value than an attractive haircut.
…the less they are founded through other values (classical distinction between intrinsic and extrinsic values)	Dignity is a higher value than amusing, which caters to dignity.
… the deeper the satisfaction connected with feeling them	A deep life satisfaction is of higher value than feeling happy while on a walk.
…the less the feeling is relative to the positioning or existence of a specific bearer of feeling or preferring	Moral values (e.g., fairness) are higher than a value such as convenience, which needs a bearer (a situation or thing that is convenient).

Figure 5 – Extrait de IEEE [2021], annexe B

Cependant, Hénin va aussi plus loin en concevant, toujours dans un scénario boîte noire, un deuxième outil, Algocate. Ce dernier prétend relier des règles symboliques proposées par l'utilisatrice ou inférées par le système à partir des données, avec des « normes » entendues également comme des règles symboliques et écrites par les concepteurs ou les auditeurs du système informatique. Hénin s'inscrit ainsi dans la lignée des travaux consacrés à l'éthique de l'intelligence artificielle qui s'efforce de formaliser le bon comportement d'un système. Mais, de l'aveu même de Vallet et Henin [2022], « [Algocate] n'était pas adapté aux contrôles et audits, car les responsables de traitement dont les systèmes d'IA seraient audités n'explicitent généralement pas les normes de la manière dont l'exige Algocate ». En effet, le recours à un algorithme qui ne peut s'expliquer qu'en boîte noire est généralement effectué lorsqu'il n'existe pas de normes (juridiques ou autres) spécifiant le comportement attendu de l'algorithme. Dans ce cas, la norme au sens de l'éthique de l'intelligence artificielle se voit formuler en principes moraux intangibles (loyauté, vigilance, etc.) que tout système devrait posséder. La quête de formalisation de principes qui devraient régir les systèmes informatiques a donné lieu, depuis une courte dizaine d'années, à un foisonnement de chartes éthiques comme celle de l'IEEE [2021] (Figure 5). Des GAFAM au Vatican, l'explicitation des valeurs morales des algorithmes semble devenue un exercice de vertu, qui déchaîne notamment la critique des philosophes du droit comme M. Hildebrandt [2020], qui leur reprochent cet abord des technologies extra-mondain et entièrement délié du droit.

Explicabilité de l'administration pour elle-même Le système d'information (SI) de la Caisse Nationale d'Allocations Familiales (CNAF) est composé de multiples applications, et l'une d'entre elles nous intéresse plus particulièrement pour contraster les deux cas précédemment évoqués : il s'agit de CRISTAL, détaillé notamment par Kounowski [2002] ou Vasutiu, Jouve et Amghar [2006]. Ce système, utilisant le langage COBOL, conçu et déployé entre 1992 et 1996, a permis d'automatiser intégralement le calcul de toutes les prestations sociales (jusqu'alors calculés par des programmes séparés) et d'obtenir en temps réel le résultat du calcul des droits personnels au moyen d'une architecture distribuée s'appuyant sur Internet. CRISTAL a ainsi fait de la CNAF une administration technologiquement très avancée, apte à absorber la complexification de l'attribution des droits sociaux caractéristique de « l'esprit gestionnaire » depuis les années 1980 [Ogien 1995]. Au fil des ans, la question du maintien de CRISTAL à jour des évolutions législatives et réglementaires des prestations sociales est apparue comme étant un enjeu crucial pour la direction des systèmes d'information de l'institution qui a conclu un accord avec le laboratoire d'informatique de l'INSA à Lyon et a financé plusieurs thèses sur le sujet [Chabbat 1997 ; Jouve 2003 ; Vasutiu 2009]. Cet axe de recherche s'appuie sur les travaux en intelligence artificielle du tournant des années 2000 portant sur l'organisation automatique de l'information grâce à des ontologies et d'autres systèmes semi-structurés permettant la recherche sémantique et l'inférence d'informations, alors vus comme une façon de faire sens de l'information disponible sur Internet [Berners-Lee, Hendler et Lassila 2001 ; Hitzler 2021].

Aussi, le dispositif technique d'explication de l'algorithme de calcul des prestations sociales gérées par la CNAF est appréhendé comme un outil interne de la direction des systèmes d'information, visant à assurer la bonne maintenance de ses programmes informatiques et la gestion des bibliothèques documentaires à destination des différents métiers de la CNAF.

La sélection de ces trois cas dans lesquels l'explicabilité d'un traitement algorithmique de l'administration a donné lieu à une réalisation permet de mettre en lumière des contraintes non plus juridiques, mais également pratiques afin d'esquisser un cheminement collectif sur les formats et les publics de la transparence algorithmique. Dans le premier cas, si l'approche par le design graphique est éminemment séduisante, elle apparaît irréaliste eu égard à la constante évolution des systèmes d'information des administrations, comme ceux calculant la taxe foncière des contribuables, et qui rencontrent déjà de lourds enjeux de maintenance. Le second cas, conçu à partir du défi représenté par les évolutions de l'intelligence artificielle et la place prise dans la recherche en informatique et le débat public par les algorithmes en « boîte noire », propose un modèle ambitieux, qui ne correspond toutefois que partiellement à la réalité des algorithmes des administrations, qui ne sont pas à proprement parler des boîtes noires dans la mesure où ils sont explicitement contraints par des règles de droit, transcrites par les DSI dans les spécifications fonctionnelles de chacun de leurs programmes. Le dernier cas apporte un éclairage sur la manière dont raisonne l'une des administrations qui utilisent le plus massivement des traitements algorithmiques en France. La CNAF conçoit la transparence algorithmique uniquement pour elle-même, pour son bon fonctionnement interne et fait primer des enjeux de sécurité des systèmes d'information et la volonté de ne pas trop renseigner d'éventuels fraudeurs sur une politique de redevabilité algorithmique vis-à-vis des citoyennes, quant bien même celle-ci est garantie par la loi informatique et libertés.

3 Enquête sur un algorithme controversé : le calcul des aides au logement

Maintenant que nous avons passé en revue les formes d'explicabilité qui concourent à la transparence algorithmique, à la fois pour les usagères et pour la bonne marche de l'administration, nous avons sélectionné un algorithme faisant l'objet d'un problème public particulièrement sonore dans l'espace public français des années 2021-2023 [Knaebel 2022; D. Minot et V. Minot 2022; Zerouala 2021] : le calcul des aides au logement.

Cet algorithme, opéré par le programme CRISTAL de la CNAF, est « entièrement spécifié par le droit » et, plus particulièrement, par le code de la construction et de l'habitation. En effet, le droit à l'aide au logement est ouvert à toute bénéficiaire remplissant les conditions légales d'accès ; son montant est déterminé à partir des caractéristiques du ménage par des règles de calcul. En raison de l'impossibilité de travailler à partir du code source CRISTAL (des fichiers COBOL envoyés par CD-ROM sans documentation technique [CADA 2019]), nous avons répliqué le code source à partir des textes de droit avec Catala [Merigoux 2023; Merigoux, Alauzen et Slimani 2023]. La réplication, complétée par des entretiens avec les rédactrices des textes dans les ministères, nous a fourni l'occasion de mettre en exergue à une autre facette de l'explicabilité algorithmique : la difficulté pour la législatrice de connaître la manière dont un programme informatique exécute le droit qu'elle édicte (ce que nous avons nommé : la frontière de l'automatisation).

3.1 Objectif de l'enquête et méthodologie

Nous avons d'abord cherché à identifier les utilisatrices potentielles d'un dispositif d'explicabilité de la distribution d'aides personnelles au logement, qui soit à la fois individualisé, détaillé et intelligible au sens de l'alinéa 2 de l'article 47 de la loi informatique et libertés. Faut-il concevoir un dispositif pour les usagères de la CAF ? Si oui, dans la lignée des dispositifs pédagogiques précédemment évoqués (Section 2.2), faut-il cibler la citoyenne concernée par ses droits, soucieuse de comprendre la raison pour laquelle elle perçoit un montant donné d'aides au logement ? Ou bien celle qui, prenant la direction du contentieux, chercherait à démontrer pour elle-même ou ses clients avoir été lésée par la CNAF et ainsi bâtir un outil de judiciarisation visant à renverser le rapport de force généré par l'opacité de l'informatique administrative ? À l'inverse, faut-il concevoir un dispositif à destination des médiatrices, des assistantes sociales et des professionnelles de l'accès aux prestations sociales dont l'activité même consiste à faire prendre conscience aux personnes de leurs droits garantis par l'État et donc à expliquer le droit ? Ou bien, faut-il directement concevoir un outil pour les agents de la CNAF, qu'ils et elles soient techniciennes-conseils dans une CAF ou informaticiennes responsables de la maintenance de CRISTAL ?

Pour répondre à ces questions liminaires, entre février à décembre 2023, nous avons pris contact avec des professionnelles en prise avec la liquidation des aides au logement : les assistantes sociales et les associations qui accompagnent les usagères dans leurs démarches, y compris contentieuses, les agents des CAF qui opèrent la liquidation, la DSI de la CNAF qui maintient le moteur de calcul des aides et un magistrat responsable de ce contentieux au tribunal administratif ($N = 8$ entretiens, avec 8 personnes). Les entretiens semi-directifs ou ethnographiques d'une durée moyenne d'une heure et demie (voir Figure 6) ont été obtenus séquentiellement, chaque rendez-vous ayant complété la liste des personnes à rencontrer. À la CAF, les interlocutrices ont été contactées en parallèle de l'étude de Lequesne-Roth, Kimri

et Legros [2021] consacrée à la dématérialisation, qui est un sujet plus vaste que celui qui faisant l'objet du présent rapport. Nous avons également revisité une partie du matériau collecté entre avril et août 2022 lorsque nous enquêtions auprès des tutelles de la CNAF, sur les modalités d'écriture du droit du logement ($N = 4$ entretiens, avec 13 personnes).

Rédactrices du droit des aides au logement
Rédactrices à la direction de l'habitat, de l'urbanisme et des paysages
Rédactrices à la direction de la sécurité sociale
Rédactrice à la direction du budget
Responsables des aides sociales à la direction du budget
Agents de la CNAF
Directeur d'une CAF
Conseiller-gestionnaire dans une CAF
Responsable d'accueil dans une CAF
Chef de projet informatique à la DSI de la CNAF
Autres professionnelles concernées par l'explication
Assistante sociale
Juriste spécialiste de l'accès aux droits sociaux
Directrice d'une association d'accès aux droits pour les travailleurs·euses immigré·e·s
Magistrat responsable du contentieux social dans un tribunal administratif

Figure 6 – Liste des personnes rencontrées

Au cours de chaque entretien, nous avons présenté le contexte de notre travail, puis nous avons orienté les enquêtées vers la description des pratiques professionnelles, en les aiguillant par des questions de mise en situation face à des demandes d'explication par une usagère, ainsi que sur le rapport au contentieux, gracieux ou judiciaire. L'analyse des propos rapportés et des documents collectés lors des entretiens ou à l'issue de ceux-ci auprès de ces professionnelles nous a permis d'appréhender la nature des problèmes qui se posent dans l'accès aux aides au logement et qui sont indissociables de l'explication du calcul. Les entretiens ont été poursuivis d'échanges par mail sur les versions intermédiaires du présent rapport ($N = 14$).

Nous présenterons les résultats de l'enquête exploratoire en suivant deux chronologies de restitution. D'abord, nous suivrons la séquence de traitement d'une décision algorithmique en analysant les opératrices des changements successifs, soit en partant de l'énonciatrice, l'usagère, qui demande à bénéficier de l'aide au logement, en passant par les médiatrices de sa demande comme l'assistance sociale, jusqu'à la DSI dont les programmes calculent le montant des droits et autorise le versement d'une somme en euro. Ensuite, nous suivrons la séquence et les potentielles actrices de contestation d'une décision administrative automatisée. Dans les deux cas, les matériaux et documents issus de l'enquête ont été sélectionnés de manière à expliciter le plus précisément possible les actrices et les circonstances dans lesquels une technologie d'explication individualisée, détaillée et intelligible de la décision algorithmique pourrait rencontrer un public et un usage.

Certains des résultats reprennent et s'adossent sur des conclusions déjà documentées par la littérature, au-delà du seul cas des aides au logement. Elles sont rappelées en vue d'insister sur l'articulation entre des problèmes souvent anciens de l'interface administrative et l'enjeu d'explicabilité des décisions algorithmiques, objet de ce rapport.

3.2 Séquence de décision administrative automatisée

L'allocataire calculée, destinatrice finale de l'explicabilité Dans le cas des aides au logement, en France, l'initiatrice d'une décision administrative automatisée est la potentielle allocataire, dont la situation personnelle sera calculée par la CNAF en vue d'accéder à sa demande, à partir du formulaire qu'elle a rempli et des pièces justificatives qu'elle a assemblées. En théorie, c'est donc elle qui est en première cheffe intéressée par l'explication de la décision automatique prise pour sa situation.

Or, les travaux de sociologie sur les usagères des services publics et la conscience du droit ont mis en exergue que la compréhension d'une interface administrative était variable selon les propriétés des groupes sociaux, leur familiarité avec les dispositifs administratifs et la manière dont l'administration elle-même les catégorise [Ewick et Silbey 1998 ; Siblot 2006 ; Weller 1999]. Et, si la compréhension des formulaires et des échanges avec les agents est précaire et inégalement répartie socialement, on peut faire l'hypothèse que la potentielle allocataire aura encore plus de mal à saisir l'enchevêtrement épais de règles, de taux et de coefficients, qui participent de la décision administrative et que les informations qui lui importent sont d'ordre pragmatique : si elle est ou non éligible à l'aide au logement, si c'était le cas, le montant qui lui sera versé, la date du versement, les pièces qu'il doit fournir pour activer ce droit et la procédure à suivre. Les travaux consacrés au non-recours ont également montré que la visibilité de la complexité administrative d'une démarche pouvait dissuader certains usagères de bénéficier de leurs droits sociaux [Deville 2023 ; Warin 2016].

Cela ne signifie en aucun cas que l'administration ne doit pas rendre de comptes sur la décision au principal intéressé. L'effort de l'explicabilité de la décision algorithmique est nécessairement orienté vers cette destinatrice finale, mais le format d'une explication réellement détaillée et la technologie d'explicabilité s'adressent à d'autres actrice que l'allocataire calculée par un traitement automatique — à commencer par les médiatrices qui remplissent traditionnellement le rôle de justification de la décision de la CNAF.

L'assistante sociale, garante de la constitution du dossier Pour les usagères moins autonomes dans la relation administrative, les assistantes sociales sont les principales médiatrices de la décision que l'administration prend à leur égard. Elles constituent des intermédiaires du droit — plus exactement, des droits — en ce sens, qu'en tant que profession non-juridique, elles mettent en œuvre au quotidien les catégories du droit et contribuent donc à la production continue de la légalité [Pélisse 2019]. Dans le cas des aides au logement, celle que nous avons rencontrée nous a expliqué que son activité était orientée vers le fait d'aider les potentielles allocataires à réunir les pièces susceptibles de former leur dossier et à remplir avec le maximum de justesse le formulaire de demande. En d'autres termes, après avoir informé les personnes sur les droits qui leur seraient potentiellement ouverts, elle traduit une situation personnelle dispersée dans une pluralité de documents dans les cases standardisées d'un formulaire et sélectionne les justificatifs pertinents. À partir de la saisie de ces données et des outils dont elle dispose (soit le simulateur d'aide au logement de la CNAF ou le simulateur multi-prestations `mes-aides.org`), elle fournit à la demandeuse une estimation de ses droits et lui dispense des renseignements sur les procédures à suivre et les autres prestations auxquelles elle est susceptible d'être éligible.

Dans la perspective de l'explicabilité algorithmique de la décision administrative du présent rapport, nous avons appris lors de cet entretien que les assistantes sociales dont elle se faisait la porte-parole ne s'intéressent que très peu à l'explication de la décision administrative ; elles considèrent que celle-ci est de la juridiction de la caisse qui traitera le dossier :

« *La CAF c'est l'État, l'État c'est le droit. Donc on ne conteste pas ce que nous dit la CAF. C'est ce qu'on nous a appris à l'école* » (entretien avec une assistante sociale, 18 février 2023). Elles ont en effet reçu une formation aux droits sociaux davantage axée sur les procédures administratives que sur la lettre du droit (en l'espèce, pour les aides au logement, le code de la construction et de l'habitation), et ont accès à une documentation professionnelle limitée. Cette assistante exemplaire par son haut niveau d'engagement professionnel nous a même expliqué « *payer [elle-même] pour m'abonner à une revue professionnelle* » (*ibid.*) et se tenir à jour des réformes. En outre, lorsque nous l'interrogeons sur les situations dans lesquelles elle a été amenée à expliquer le montant qu'elle a simulé ou l'écart entre le montant qui avait été annoncé et celui qui a été perçu à une allocataire curieuse, circonspecte, abattue ou remontée, elle explique qu'elle sait composer avec l'imprécision : « *c'est une simulation, on dit que le vrai montant sera différent* » (*ibid.*). Même dans les cas où elle s'est aperçue que la CAF s'était trompée, elle a signalé l'erreur, mais n'a pas cherché à entrer dans les plis de l'attribution et du calcul des droits. On apprend dans l'entretien avec cette assistante particulièrement diligente que « *la CAF refusait de donner ses aides à un Polonais hébergé dans un foyer parce qu'il n'avait pas de titre de séjour. Mais la Pologne est dans l'Union Européenne, son passeport suffisait ! J'ai envoyé un courrier recommandé à la CAF mais n'ai pas eu de réponse, j'ai alors laissé tomber* » (*ibid.*).

En conséquence, même si les assistantes sociales se trouvent de facto régulièrement en situation d'expliquer ou de justifier individuellement une décision administration sur l'attribution ou le calcul des droits sociaux, elles renvoient les allocataires vers la CAF et estiment que leur « vrai boulot » [Bidet 2010] réside dans l'information sur les droits, puis la constitution du dossier et la justesse du remplissage des formulaires. Il faut donc pousser la porte de l'administration et chercher plus loin dans la circulation de la décision administrative qui pourrait mettre en œuvre de manière adéquate et circonstanciée l'exigence d'explicabilité algorithmique définie par le droit.

Les conseillères de services à l'usagère des CAF, appuis de la constitution du dossier et relais ponctuels Les agents responsables de l'accueil dans les CAF — appelées conseillères de service à l'usagère — poursuivent le même objectif d'information sur les droits et d'accompagnement que les assistantes sociales. Elles listent les pièces à apporter pour formuler une demande d'aide et disposent pour cela à la fois une connaissance des principes généraux du droit, des principaux dispositifs mis en œuvre par la CNAF et savent identifier et gérer différents profils types d'usagères. Elles se distinguent toutefois des assistantes sociales dans la mesure où, du fait du cadre de la rencontre avec les usagères — dans l'espace d'accueil de l'agence avec une confidentialité limitée, où souvent beaucoup de personnes attendent en même temps — elles ne nouent pas de relations personnalisées avec les potentielles allocataires et ne les accompagnent que pour une demande ponctuelle auprès de la CAF.

Un déplacement dans une agence nous a appris que l'accueil est supposé résoudre les difficultés les plus courantes des potentielles allocataires (création de compte en ligne, téléchargement de documents, etc.) et « *se concentre sur l'établissement du dossier de l'allocataire* » (entretien avec une responsable d'accueil de CAF, 7 juin 2023). Les « *conseillers de service à l'usager (CSU) n'ont pas le même niveau de formation technique et juridique que les conseillers-gestionnaires [responsables du traitement des dossiers] avec lesquels on peut prendre rendez-vous en présentiel, par téléphone ou en visioconférence* » (correspondance avec une responsable d'accueil de CAF, 10 octobre 2023). Elles ne sont donc, pas plus que les assistantes sociales, en mesure de fournir des explications, susceptibles de remplir l'exigence légale. Toutefois, leur position interne aux agences rend possibles certains aménage-

ments. Par exemple, la prise en compte des critiques du choix initial de faire de ces agents les strictes accompagnatrices des démarches en ligne leur ont permis d'accéder à des formations sur les droits sociaux : « *avant ils ne pouvaient qu'accompagner sur les démarches en ligne et donner des informations d'ordre général sur les aides et n'avaient aucune bille pour répondre aux questions précises des gens* » (entretien avec une responsable d'accueil de CAF, 7 juin 2023). En outre, si l'accès aux conseillères-gestionnaires, seules habilitées à liquider les prestations et donc à déclencher les versements aux allocataires, est délibérément filtré, cette règle laisse place à des ajustements pratiques, en fonction de la complexité ou de l'urgence de la situation : « *quand on a un cas de violences conjugales au comptoir, on le reçoit immédiatement avec un conseiller-gestionnaire et on boucle le dossier* » (*ibid.*).

Ainsi, même si les conseillères de service à l'usagère dans les agences sont le visage de la CNAF pour le public et se trouvent, comme les assistantes sociales, en situation d'expliquer les décisions d'attribution d'aide au logement aux personnes concernées, leurs principales tâches consistent à informer sur les droits, puis à aider les potentielles allocataires à remplir leur demande et à fournir les pièces justificatives. Fort de leur proximité dans la CAF avec les conseillères-gestionnaires, elles peuvent ponctuellement réparer des situations, mais, depuis leur poste de travail, n'ont pas l'équipement technique ni la formation nécessaire pour expliquer de manière individualisée, détaillée et intelligible la décision automatique de l'administration. Il faut donc se déplacer plus loin encore et suivre le cheminement de la décision dans l'administration pour identifier les potentielles actrices de l'explicabilité, telle que définie par l'alinéa 2 de l'article 47 de la loi informatique et libertés.

Les conseillères-gestionnaires, les nécessités de la liquidation en masse et les justifications en marge Les conseillères-gestionnaires des CAF reçoivent certaines catégories de publics en rendez-vous, valident les données saisies par la demandeuse (directement ou par l'intermédiaire d'une assistante sociale), vérifient les pièces du dossier, déclenchent la liquidation des droits et notifient la bénéficiaire. Elles sont donc elles aussi des intermédiaires du droit, au sens où elles mettent en pratique les catégories du droit et disposent d'un accès à une vaste documentation interne sur les droits et les procédures de la CNAF, mais n'exercent pas une profession juridique au sens strict [Pélisse 2019]. Dans leurs termes, « *je ne connais pas le droit. Tout est résumé sous la forme de fiches sur @Doc [outil interne de documentation]. Il n'y a que les services centraux à la CNAF qui se réfèrent au droit* » (entretien avec un conseiller-gestionnaire, 7 juin 2023). Plus encore que les assistantes sociales et les agents d'accueil qui ne contribuent qu'à une partie des dossiers de demande d'aide au logement, il est important d'avoir à l'esprit que les conseillères-gestionnaires s'occupent d'une masse de demandes. Il faut également noter qu'il existe au sein du groupe professionnel une division du travail entre des pôles, permettant aux agents de se familiariser avec une partie des 25 prestations délivrées par la CNAF.

Lorsque nous interrogeons un conseiller-gestionnaire sur l'explication des décisions en matière d'aide au logement, il estime que c'est bien à lui que revient la tâche d'expliquer la décision administrative aux allocataires concernées, soit lors du rendez-vous, soit en écrivant un courrier ou message dans l'espace personnel sur `caf.fr`. Toutefois, ces dernières sont prudentes dans les justifications qu'elles fournissent et reconnaissent la complexité de cette tâche : « *il faut être vigilant quand on donne des explications avec les réformes* » (entretien avec un conseiller-gestionnaire, 7 juin 2023), et ce même si la CNAF tient à jour la documentation @Doc (circulaires, notices, fiches techniques) sur les évolutions réglementaires. Plus précisément, elles estiment être en mesure de saisir les variations d'éligibilité et de montants des prestations en fonction des données d'entrée, notamment dans

les cas délicats d'indus. Autrement dit, par un jeu de déductions s'appuyant sur leur expérience professionnelle et les 18 mois de leur formation initiale, elles sont en mesure d'expliquer, relativement, la décision prise par le système d'information pour un cas particulier. Lorsqu'un dossier déposé sur `caf.fr` est traité, « *la notification émise par le système d'information est généraliste et peut manquer de précision. Cependant, lorsqu'un gestionnaire régularise lui-même un dossier, la notification est rédigée à la main par ce dernier, des explications supplémentaires peuvent alors être apportées à l'allocataire* » (correspondance avec un conseiller-gestionnaire, 30 octobre 2023). Dans les termes des agents, il s'agit le plus souvent d'« *un message de quelques lignes que je laisse sur leur dossier sur `caf.fr`* » (entretien avec un conseiller-gestionnaire, 7 juin 2023). Le temps consacré à la rédaction de cette explication doit être entendu dans le contexte de la corbeille typique d'une conseillère-gestionnaire contenant « *4 231 dossiers à traiter* » (*ibid.*) qui ne sont qu'une partie des « *26 000 dossiers à traiter pour toute la CAF [départementale]* » et représentent « *l'équivalent de 3 jours de travail si la CAF ne reçoit plus aucun courrier (ni papier, ni dématerialisé)* » (correspondance avec un conseiller-gestionnaire, 30 octobre 2023). S'« *il est évident que le temps de l'explication est compté dans notre temps de travail* » (*ibid.*), l'évaluation professionnelle des conseillères-gestionnaires se fait entre autres sur la base de la « *performance [...] comparée à une moyenne collective* » (*ibid.*), la performance étant ici entendue comme le nombre de dossiers traités par agent. La présence de cette masse de dossiers et d'évaluation de la performance ne signifie pas que les agents n'aient pas un souci de la personne et n'essaient pas de procéder à des accommodements au bénéfice des allocataires : « *Ainsi, une priorisation peut être donnée aux populations les plus vulnérables telles que les bénéficiaires du RSA et du Handicap. En outre, le traitement des pièces, quel que soit le segment de population, n'obéit pas à une règle de traitement par antériorité. Des pièces prioritaires sont identifiées, telles que les pièces permettant de lever une suspension, d'ouvrir ou de maintenir des droits. Enfin, les gestionnaires peuvent suivre des dossiers chaque jour, ils sont informés des réponses données par les allocataires suite à des appels de pièces, suivent des échéances qu'ils ont positionnées sur des dossiers, reprennent des dossiers qu'ils ont mis en attente* » (*ibid.*).

Il reste, outre l'enjeu d'inégalité d'accès documenté par des rapports administratifs et des enquêtes de sciences sociales (récemment : [Deville 2023]), que l'explication des résultats de liquidation des droits semble, dans la pratique, inégalitaire dans la mesure où elle varie selon si le dossier a été régularisé automatiquement ou par une agent, selon la fréquence du cas, le temps et les ressources documentaires consultées par ce dernier. Pour cette raison, l'explication telle que mise en œuvre par les conseillers-gestionnaires de la CAF ne semble pas en mesure de se placer à la hauteur de l'ambition d'individualisation, de détail et d'intelligibilité fixée par les textes et cela n'est en aucun cas la conséquence d'un défaut de formation, de temps ou de moralité des agents. De la même manière que tous les dossiers sont liquidés par un même programme — et même si les questions de liquidation et d'explication n'ont jusqu'alors pas été pensées ensemble —, il s'agit-là d'un enjeu d'infrastructure administrative d'explication de la décision automatique et de traitement commun à tous les usagères. En conséquence, les conseillers-gestionnaires pourraient être les usagères ordinaires d'une technologie d'explicabilité des décisions automatiques, qui leur apporterait un supplément d'assurance et de précision dans leur travail de justification de l'attribution et de calcul administratif au bénéfice de la personne concernée.

Le Pôle d'appui technique, ressource pour la liquidation des cas particuliers et relai vers la DSI Avant de venir au système d'information de la CNAF en tant que tel, il faut noter que

chaque département dispose d'un pôle d'appui technique. Les pôles remplissent, sur un petit nombre de cas ou pour des configurations nouvelles, une fonction d'explication des règles de calcul, à destination des conseillers-gestionnaires et, si ces derniers les répercutent dans les notifications, à l'attention de la bénéficiaire final de l'allocation. En d'autres termes, dans la configuration actuelle de l'organisation de la CNAF, les agents de ces pôles constituent des actrices clés de l'explication de la décision automatique.

Nous avons appris au cours de l'enquête que les conseillers-gestionnaires pouvaient recourir à outil, SAXO, pour signaler leurs difficultés dans la liquidation de certains dossiers aux agents du pôle d'appui technique. Ces dernières disposent de davantage de temps pour consulter la documentation interne nécessaire au traitement de ces cas particuliers et peuvent réaliser manuellement une liquidation en cas de problème interne au système d'information. Dans ce cas, elles signalent elles-mêmes le problème à la DSI. Or, la Cour des Comptes [2023] a relevé que les limites de cette organisation : « les agents de la branche effectuent toujours les actes de liquidation les plus complexes sans que le système d'information leur procure une assistance suffisante, ce qui accentue les risques d'erreurs inhérents aux traitements manuels [...] les contrôles effectués en 2022 par les directions comptables et financières portent sur une fraction réduite des liquidations et font apparaître des taux d'erreurs significatifs, concernant notamment les ressources prises en compte ». Comme les conseillères-gestionnaires, les agents du pôle d'appui technique sont des actrices de l'explication de la décision automatique de la CNAF, à la différence qu'elles ne reçoivent pas les allocataires et qu'elles rédigent des explications seulement sur un petit nombre de cas particulièrement complexes ou nouveaux. Aussi, leurs explications peuvent également s'avérer plus ou moins étayées selon l'état de la documentation, le temps dont elles disposent et l'état de fonctionnement du système d'information. Ces agents pourraient également être des usagères d'une technologie d'explicabilité des décisions automatiques au sens de la Section 2.1 et, plus encore que les conseillères-gestionnaires, pourraient remonter, de la même manière qu'elles le font actuellement avec le système de liquidation, les problèmes qu'elles détectent dans les explications. La mise en place de cette technologie constituerait aussi un gain de précision dans leur travail et leur fournirait un degré de compréhension plus avancé du fonctionnement de la CNAF.

La direction des systèmes d'information de la CNAF, clé de voûte de la décision automatique et boîte noire de l'explication La direction des systèmes d'information de la CNAF s'occupe, entre autres choses, de maintenir et de corriger le programme de liquidations des aides et des prestations sociales et de tenir à jour le système de documentation des évolutions des droits sociaux, qui permet autant aux programmeuses de spécifier les paramètres de calcul du programme, que de guider les conseillers-gestionnaires et les techniciens du pôle d'appui dans l'explication de la liquidation des droits.

Depuis 1996, le système d'information mis en place par la CNAF pour liquider l'ensemble des prestations familiales s'appelle CRISTAL. En 2019, CRISTAL a été complété d'un système d'information conçu par le prestataire de services informatiques Oracle spécifiquement pour liquider les aides au logement ; Oracle Policy Automation n'a jamais été considéré suffisamment fiable pour remplacer CRISTAL entièrement de sorte, qu'en 2023, les deux programmes calculent les aides au logement des allocataires. Les paramètres de CRISTAL et Oracle Policy Automation correspondent à des spécifications légales et réglementaires. Or, l'interaction entre les prestations donne lieu à une combinatoire de règles d'une grande complexité : « *les aides au logement sont des prestations simples du point de vue du juriste, mais ce qui devient très compliqué ce sont les interactions avec toutes les autres prestations* » (entre-

tien avec un chef de projet informatique à la CNAF, 14 juin 2023), par exemple, parce que l'ouverture du droit à l'allocation de logement sociale suppose que l'allocataire ait préalablement été déclaré inéligible à l'allocation de logement familiale et à l'aide personnalisée au logement. Autrement dit, l'explication du calcul des droits par le système d'information est particulièrement complexe du fait de l'interdépendance des aides. Elle s'avère également complexe en raison de l'absence de traçabilité explicite entre les règles de droit encadrant une prestation donnée et ces deux programmes : « *le problème de fond c'est le lien entre les règles et la documentation réglementaire et le code* » (*ibid.*), soit ce que l'informatique juridique nomme l'isomorphisme [Bench-Capon et Coenen 1992] et qui suppose un outillage d'ingénierie logicielle sophistiqué dont la CNAF ne dispose pas. Par conséquent, la CNAF est organisée de manière à ce que des juristes de la direction des politiques familiales et sociales rédigent, à partir du droit en vigueur, une documentation de référence. Ensuite, des analystes déclinent à partir de cette documentation juridique des documentations spécialisées : l'une servira de spécification aux systèmes d'information, une autre à l'outil @Doc pour les conseillers-gestionnaires et une troisième sera destinée à la formation initiale. Puis, les programmeuses de chacun des deux systèmes d'information, Oracle Policy Automation et CRISTAL, s'appuient sur la documentation des spécifications fonctionnelles pour mettre à jour le programme suivant les évolutions du droit. Cette division du travail donne lieu à une organisation dite « en V » produisant un pipeline qui impose une contrainte immédiate de séquentialisation des changements : plutôt que de pousser deux changements à la fois dans toutes les étapes de traduction, on les pousse l'un après l'autre après toutes les étapes dans l'ordre. Et, il faut imaginer qu'il s'agit là d'une organisation proprement industrielle : « *on a des centaines de personnes qui sont chargées de mettre à jour la documentation en permanence* » (*ibid.*). Or, la lecture de la réglementation pour les seules aides au logement nous a permis de faire l'expérience [Merigoux 2023 ; Merigoux, Alauzen et Slimani 2023] d'une difficulté supplémentaire pour l'explicabilité des décisions automatiques prises par la CNAF : « *la grande versatilité et la modification en continu de cette réglementation* » (*ibid.*). En effet, la tâche de mise à jour implique des tâches d'analyse juridique et de programmation, qui représentent un travail intellectuel qui résiste à la division des tâches et à l'automatisation [2].

Les programmes de la direction des systèmes d'information, CRISTAL et Oracle Policy Automation, constituent l'instrument de liquidation automatisée des aides au logement et la clé de voûte de l'explication de la décision administrative d'attribution et de calcul de l'aide. Or, la mise en place d'un dispositif d'explicabilité à la DSI est aujourd'hui triplement contrainte par la complexité des règles, par les technologies de collaboration et de calcul et par l'organisation générale du travail à la CNAF. Le SI se présente alors davantage comme une boîte noire que comme outil d'élucidation de la décision automatisée à destination des intermédiaires du traitement et des personnes concernées. Cela revient à faire *de facto* reposer la charge de l'explication de la décision automatisée sur les déductions précédemment évoquées des conseillères-gestionnaires soutenus par le pôle d'appui technique.

Nous avons utilisé les matériaux de l'enquête exploratoire pour retracer la séquence de la décision administrative automatisée dans le cas des aides au logement et identifier les destinataires et les usagères d'une technologie d'explication du traitement automatique au sens du droit, soit individualisée, détaillée et intelligible. Nous avons vu que si la demandeuse de l'aide était la première concernée, elle était surtout une destinatrice finale, du fait de la complexité des règles de calcul qui lui sont appliquées. Il en va de même pour les média-

2. Il s'agit là d'une problématique récurrente des organisations, et à ce titre a fait l'objet de nombreux débats, depuis les universités américaines des années 1970 (par exemple [Brooks 1974], jusqu'au cabinet de conseil McKinsey en 2023 [Gnanasambandam et al. 2023]).

trices de la demande, assistantes sociales ou personnel d'accueil dans les CAF, qui, dans la division du travail, ne sont pas tenues responsables de l'explication de la décision de l'administration, mais de l'information sur les droits et du dépôt de la demande. L'explication est en revanche du ressort des conseillères-gestionnaires, qui, soutenus par un pôle d'appui technique, reçoivent les allocataires et rédigent de courtes explications des résultats de la liquidation des droits au nom de la CNAF. Or, ces dernières travaillent aujourd'hui à la main, par rapprochement de situations et à partir de déductions, ce qui génère une qualité d'explication nécessairement variable et peu exhaustive. En conséquence, en tant qu'opératrices de la liquidation des dossiers, les conseillères-gestionnaires et les agents du pôle d'appui technique pourraient être les usagères d'une technologie d'explicabilité à destination des allocataires calculées, qui aurait été conçue à la DSI, directement à partir des paramètres du calcul et de sa spécification en droit. En outre, on peut avancer l'hypothèse selon laquelle le fait de disposer d'une telle technologie faciliterait aussi la mise à jour en fonction des évolutions du droit, les tests et la correction des bugs et permettrait donc un meilleur contrôle de l'administration par elle-même. Pour prolonger l'exploration des usages de l'explicabilité des décisions administratives, il faut maintenant passer en revue une autre séquence, qui peut suivre dans certains cas le traitement d'une demande d'aide au logement : la contestation de la décision de l'administration.

3.3 Séquence de contestation des décisions automatiques

Une partie des décisions administratives sont contestées par les personnes concernées. Dans le cas des aides au logement, l'essentiel de la contestation relayée dans l'espace médiatique [Gauvin, Lerch et Krouk 2022], concerne les indus et trop-perçus. Il s'agit là d'un phénomène massif : « *on a 4 millions d'indus par an* » (entretien avec le directeur d'une CAF, 7 juin 2023), qui désorganise en partie le travail décrit dans la séquence ci-dessus, peut engendrer des situations délicates pour les agents comme les allocataires, voire nourrir de la violence à l'encontre des agents d'interface — de l'assistante sociale, à la conseillère-gestionnaire, en passant par les conseillères de service à l'usagère. Nous allons donc continuer à déplier les données collectées au cours de notre enquête pour saisir, cette fois, à partir d'une séquence de contestation, les destinataires et les usagères potentiels d'une technologie d'explicabilité de la décision administrative à la hauteur des exigences du droit que nous avons décrites dans la Section 2.1.

L'allocataire calculée, émettrice de la contestation L'allocataire qui s'estime lésée peut contester, seule ou accompagnée d'une avocate ou d'une association, la décision concernant ses droits aux aides au logement par l'intermédiaire de la commission des recours amiables ou du service de la médiation de la CAF ou peut judiciariser sa situation en faisant le choix du contentieux devant le tribunal administratif. C'est sur elle (et ses éventuelles représentantes : avocates ou association d'accès au droit) que repose la charge de la preuve.

En pratique, ces instances sont saisies lorsque « *les techniciens CAF ne savent pas répondre [aux demandes de justification de l'allocataire qui s'estime lésé], donc oui, il [lui] faut saisir la commission de recours amiable et le juge [pour obtenir une interprétation des règles de droit au cas d'espèce]* » (entretien avec une juriste spécialiste de l'accès aux droits sociaux, 6 mars 2023). Au tribunal administratif de Paris, en 2022, l'aide au logement représente un volume de 148 dossiers (soit 8% pour le contentieux social) [3] et les principales configurations soumises par les allocataires sont « *des situations portant sur l'étendue des droits pour le*

3. Source : demande au greffe du tribunal administratif de Paris en décembre 2023.

passé et l'avenir », des « *contestations de la notification d'indus ou d'une contrainte à l'initiative de la CAF* », ou « *une demande de remise gracieuse* » (entretien avec un magistrat, 14 décembre 2023) ; la première configuration étant la « *moins fréquente* », mais la plus complexe pour l'allocataire qui doit être apte à calculer elle-même ses droits et pour la magistrate dans la mesure où cette dernière doit « *sérier les périodes d'éligibilité et donc demander des pièces justificatives au requérant pour reconstituer ses droits, puis reconstituer le calcul* » (*ibid.*). C'est dans ces situations là que les parties et la magistrate qui doivent tous recalculer à la main les montants d'aide font l'expérience que le calcul proposé par la CAF initialement, celui de l'avocate basé sur le guide de calcul édité par le bureau des aides personnelles au logement [Ministère chargé de la ville et du logement 2023], celui de magistrate qui est reparti de la lettre du droit et du mémoire en défense de la CNAF diffèrent amplement (*ibid.*).

Aussi, pour l'allocataire calculée, la mise en place d'outils d'explication de la décision automatique peut être un appui conventionnel pour la contestation grâce auquel elle va comprendre, dans la complexité des étapes de calcul, la manière dont sa situation personnelle a été qualifiée par la CAF et les ressources qui ont été prises en compte . Plus précisément, l'explication de la décision peut aider l'allocataire à administrer une partie de la charge de la preuve de la situation et à préparer les pièces justificatives nécessaires à la construction de l'argumentation, en fait et en droit, sans quoi le contentieux se trouve entravé par la difficulté (pour les parties comme pour la magistrate) à reconstituer le montant des droits. Il faut noter que la mise en place de tels outils aura un effet sur la nature du contentieux en matière de droit au logement dans la mesure où celui-ci deviendra certainement plus technique et obligera la juge à entrer à son tour dans les plis du calcul des droits.

Les associations d'accès au droit, militer pour le droit contre l'État À l'inverse des assistantes sociales qui sont animées par l'idée selon laquelle « *L'État c'est le droit* » et qui se faisant laissent à l'État et à ses différentes instances le monopole d'énonciation de la légalité et de la répression des illégalismes, les associations d'accès au droit qui accompagnent une partie des allocataires dans leurs démarches administratives sont porteuses, au nom du droit, d'une critique du fonctionnement de l'État. En cherchant ainsi à éprouver la conformité juridique de l'activité étatique et à mettre en lumière ses irrégularités, elles prolongent les mouvements de retournement du droit contre l'État qui se sont développés tout au long du XXe siècle (*speaking law to power* [Abel 1998]) : suffragettes, objectrices de conscience, militantes des droits civiques, lanceuses d'alerte, etc.

Les associations de l'accès au droit agissent sur des cas individuels en suivant des allocataires ou de potentielles allocataires dans leurs démarches (par exemple, de contestation des indus ou des trop-perçus réclamés par la CNAF), se forgeant, au fil des dossiers, une expertise juridique et administrative sur le droit du logement et s'engageant par et pour le respect du droit contre l'État. Nous avons par exemple rencontré une association spécialisée dans l'accompagnement des personnes immigrées, consciente de la technicité du contentieux en matière d'allocations au logement et des difficultés pour une structure subventionnée d'y prendre part : « *notre association fait de l'accompagnement socio-administratif et juridique pour permettre aux milieux populaires, en particulier immigrés, d'accéder à leurs droits et pour cela nous allons si nécessaire jusqu'aux contentieux (entre autres contre les CAF). La plupart des autres structures qui font de l'"accès aux droits" ne vont pas aussi loin et n'accompagnent pas aussi systématiquement, car la tâche est chronophage, technique et quasiment plus subventionnée alors que cela reste officiellement un objectif des politiques publiques* » (entretien avec la directrice d'une association, 10 mars 2023). La responsable de cette structure relève la fragilité de cet engagement, qui est à la fois technique et chrono-

phage lorsqu'il s'agit de recomposer le raisonnement de la CNAF sur un dossier, et qui se trouve dans une situation de dépendance vis-à-vis de subventions attribuées au prorata du nombre de personnes accompagnées : « *le subventionnement est purement quantitatif et incite les structures à diminuer le temps consacré à chaque allocataire, donc à ne pas engager de contentieux pour rétablir des droits de plus en plus souvent bafoués. Ainsi nous avons perdu plus de la moitié de nos subventions depuis 2017, et la situation n'était plus économiquement viable. Pour éviter le dépôt de bilan et honorer nos engagements jusqu'au bout, nous avons dû prendre avec le conseil d'administration la difficile décision d'organiser la cessation d'activité de notre association* » (*ibid.*). Le temps et la technicité requis pour préparer la contestation des décisions administratives ne sont pas les seules variables. L'association que nous avons rencontrée sait que son expertise est limitée par le manque de transparence de la CNAF quant à la déclinaison locale des règles de droit : « *Même avec des connaissances précises de la loi, on arrive pas toujours à expliquer pourquoi il y a une suspension de prestation ou une baisse de montant lorsque les décisions des CAF ne sont pas motivées en droit, ce qui est rarement le cas* » (entretien avec une juriste spécialiste de l'accès aux droits sociaux, 6 mars 2023). Pour élucider ces situations, « *on cherche à avoir les circulaires ou instructions internes* » (*ibid.*), mais certaines d'entre elles font l'objet d'un refus de publication de la part de la CNAF [2023] : « *il serait incompréhensible d'attendre [...] la publication de la moindre instruction donnée [par la CNAF] à son réseau, dont la plupart n'ont aucun impact sur l'application des règles de droit, et ne portent que sur des dispositions de gestion interne, comme par exemple les modalités de traitement d'une situation dans [son] système d'information* ». Cette opacité de la déclinaison administrative des textes de droit emporte des effets sur la nature de l'argumentation de l'association : « *il y a bien des dossiers où on peut argumenter sur l'éligibilité aux APL, de l'éligibilité des gens, mais le calcul du montant on ne dispose pas des éléments pour le faire* » (entretien avec la directrice d'une association d'accès aux droits, 10 mars 2023).

En conséquence, les associations d'accès au droit pourraient bénéficier à plusieurs égards de la mise en œuvre d'un dispositif d'explicabilité des décisions administratives. Elles disposeraient à la fois de davantage de prises pour construire leur argumentaire et administrer la charge de la preuve pour le compte des personnes calculées. En même temps, elles renforceraient leur expertise sur la mise en application du droit par l'administration et pourraient ajuster ou déplacer la critique de l'État par le droit qui est au cœur de leur engagement. Les associations d'accès au droit dont il a été question jusqu'ici font partie d'un ensemble d'actrices qui exercent de manière plus ou moins systématique et formelle une vigilance sur l'activité administrative. Il est maintenant temps de passer en revue cet écosystème de vigilance et de contrôle à la fois, interne et externe, à l'administration qui serait susceptible de bénéficier dans l'accomplissement ou dans le renouvellement de leurs missions, d'outils d'explicabilité de la décision administrative conformes à l'alinéa 2 de l'article 47 de la loi informatique et libertés.

Les potentielles actrices d'un contrôle interne et externe de la décision administrative automatisée Au cours de notre enquête par entretiens et de la collecte de documents portant sur l'attribution et le calcul des aides au logement, nous avons identifié deux types d'actrices susceptibles de bénéficier, indirectement, de l'explication individualisée, détaillée et intelligible d'une décision administrative prise sur le fondement d'un traitement automatique de données. Nous distinguons les actrices internes à l'administration, qui sont, dans le cas de l'attribution des aides au logement, d'abord les tutelles administratives de la CNAF (direction de la sécurité sociale, direction du budget, direction de l'habitat, de l'urbanisme et

des paysages), puis les actrices de la transparence algorithmique (la mission Etalab et la CNIL), mais aussi les corps d'inspection (inspection des affaires sociales et inspection générale des finances) et la Cour des comptes qui, nous l'avons lu, n'a pas certifié les comptes de la CNAF en 2022 en raison des problèmes de liquidation des droits [Cour des Comptes 2023]. À côté de ces actrices, nous avons cerné les contours d'un réseau hétérogène situé en dehors de l'administration et concerné par ou impliqué à des degrés divers dans l'explicabilité algorithmique des décisions administratives. Il y a d'abord le Parlement qui vote le projet de loi de financement de la Sécurité sociale et a dans ses attributions l'évaluation de l'application des lois, qui serait facilitée par la mise en place de l'explicabilité, il y a aussi des associations dont les associations d'accès aux droits précédemment évoquées et les associations de préservation des libertés numériques (principalement : La Quadrature du Net ou Le mouton numérique), auxquelles s'ajoutent certaines petites entreprises, associations ou coopératives qui conseillent les actrices publiques sur le sujet, dont Datactivist. En plus de ces actrices institutionnelles, des journalistes qu'elles soient spécialisées dans les questions numériques comme Hubert Guillaud, rédacteur du blog *Internet Actu*, ou le pôle enquête des Décodeurs du journal *Le Monde*, qui a mené une récente enquête sur les algorithmes de la CNAF [Romain et al. 2023] et certaines chercheuses, travaillant en informatique, en droit ou en sciences sociales — dont les autrices de ce rapport — pourraient bénéficier, dans leurs travaux, des technologies d'explicabilité des décisions administratives et relayer ces enjeux vers les citoyennes.

Pour l'heure, les actrices du contrôle interne de l'administration ont une connaissance limitée de l'enjeu d'explicabilité administrative et de l'état de sa mise en application. Par exemple, lors d'entretiens avec les tutelles de la CNAF, toutes les rédactrices des textes encadrant les allocations au logement ont mis en avant la difficulté à saisir ce que faisait effectivement la CNAF des règles de droit. Par exemple, une responsable de la direction du budget mettant à jour un tableau de 114 valeurs modifiées par un décret [Merigoux 2022] et un arrêté pris à l'été 2022, se demande dans quelle temporalité et sous quelle modalité la CNAF, qui voit ses dépenses informatiques croître, applique effectivement les changements de barème (entretien avec une administratrice de la direction du budget, 19 août 2022), tandis que la direction de la sécurité sociale, explique que « *la DSI de la CNAF a toujours été une boîte noire. On n'y a pas accès* » (entretien à la direction de la Sécurité sociale, 6 juin 2022). Les agents du bureau des aides au logement ont une compréhension un peu plus précise dans la mesure où elles savent que « *les développements sont priorisés à la CNAF, en fonction de toutes les mises à jour à faire [et] qu'il y a donc un délai entre le droit et le code, qui est connu et qui est géré de manière interne à la DSI de la CNAF* » (entretien à la direction de l'habitat, de l'urbanisme et des paysages, 23 mai 2022). Autrement dit, les responsables du contrôle interne à l'administration, fixant non seulement les textes, mais aussi les conventions d'objectifs et de moyens de la CNAF doivent déléguer le contrôle à la DSI de la CNAF, qui bénéficie alors d'une large autonomie dans la mise en application des développements nécessaire à l'explication des décisions automatiques. En outre, nous l'avons déjà lu, les actrices externes à l'administration se heurtent, elles aussi, à l'indisponibilité de la documentation technique produite à la CNAF et même la juge n'est pas en mesure de reconstituer le calcul, alors même qu'il s'agit d'un objectif défini dans la jurisprudence de référence du Conseil d'État :

> « il appartient au juge administratif, eu égard tant à la finalité de son intervention dans la reconnaissance du droit à cette allocation ou à cette aide qu'à sa qualité de juge de plein contentieux, non de se prononcer sur les éventuels vices propres de la décision attaquée, mais d'examiner les droits de l'intéressé sur lesquels l'administration s'est prononcée, en tenant compte de l'ensemble des circonstances

> de fait qui résultent de l'instruction et, notamment, du dossier qui lui est communiqué en application de l'article R. 772-8 du code de justice administrative. Au vu de ces éléments, il appartient au juge administratif d'annuler ou de réformer, s'il y a lieu, cette décision en fixant alors lui-même les droits de l'intéressé, pour la période en litige, à la date à laquelle il statue ou, s'il ne peut y procéder, de renvoyer l'intéressé devant l'administration afin qu'elle procède à cette fixation sur la base des motifs de son jugement » [Conseil d'État 2012].

En conséquence, la mise en place d'outils d'explication individualisée, détaillée et intelligible des décisions administratives automatisées prises par la CNAF rencontrerait l'intérêt d'une diversité d'actrices lors des séquences dans lesquelles l'allocataire calculée conteste la décision. Plus précisément, cela bénéficierait à la fois aux responsables du contrôle interne de la CNAF dans leurs missions (d'évolution de la réglementation, de fixation des contrats d'objectifs et de moyens ou encore de certification des comptes), mais aussi aux actrices du contrôle externe, à commencer par la juge administratif.

L'enquête exploratoire menée sur le cas des aides au logement nous a permis de mettre en lumière deux potentiels contextes d'usage de l'explicabilité de la décision administrative automatique au sens de l'alinéa 2 de l'article 47 de la loi informatique et liberté, qui a fait l'objet de la première partie du rapport. Premièrement, l'explication de la décision s'adresse à l'allocataire objet du calcul des droits et aux médiatrices qui peuvent se retrouver en situation de lui traduire la décision (assistantes sociales, conseillers de service à l'usagère de la CAF). Les conseillères-gestionnaires et les pôles d'appui technique des CAF seraient quant à elles les principales utilisatrices d'une technologie visant à expliquer au sens du droit la décision, dans la mesure où elles remplissent déjà, mais partiellement, cette fonction dans le cadre de leurs missions. Toutefois pour que l'explicabilité ne soit pas une fiction pédagogique, mais un réel traitement du cas de l'usagère, une telle technologie doit être arrimée au système d'information qui calcule les prestations sociales et doit donc émaner de la DSI de la CNAF. Deuxièmement, l'explicabilité peut également s'adresser à l'allocataire qui souhaite contester la décision administrative. Dans ce cas, la mise à disposition d'une explication au sens du droit serait non seulement bénéfique à l'allocataire calculée et à ses représentantes (association d'accès au droit ou avocate), mais permettrait aussi à ces dernières de faire évoluer leurs argumentaires grâce à une connaissance plus claire du fonctionnement de la CNAF. Ainsi, des actrices administratifs (tutelles de la CNAF, missions d'inspection, Etalab) ou extérieurs à l'administration (législatrices, journalistes, chercheuses, associations d'accès aux droits ou de protection des libertés numériques) pourraient exercer un contrôle à la fois interne et externe sur les modalités de prise de décision administrative, permettant ainsi d'étendre le circuit démocratique au-delà de ce qui a été délibéré et voté au Parlement et en direction de la pratique d'administration des citoyennes.

Partant de cette réflexion sur le cas des aides au logement, nous avons conçu trois prototypes de technologies d'explication des décisions administratives conformes à la triple exigence d'individualisation, de détail et de l'intelligibilité et susceptibles de rencontrer les usages que nous avons identifiés ci-dessus. Nous les présentons dans la section suivante.

4 Prototypage : génération automatique d'explications des décisions administratives

Après avoir exploré l'état de l'art en Section 2 et enquêté sur les potentiels usages et usagères d'outils d'explicabilité de décisions algorithmiques en Section 3, nous proposons de contribuer à la formulation d'une technologie d'explication automatisée des décisions administratives automatisées conforme à l'exigence d'individualisation, de détail et d'intelligibilité. Pour cela, nous continuerons à prendre appui sur le cas des aides au logement. Notre réflexion est présentée suivant une logique d'exposition croissante du caractère détaillée de la décision administrative automatique. Pour chaque prototype, nous explicitons la méthodologie d'explication employée et discutons des avantages et inconvénients de l'explication produite.

Contrairement aux travaux du courant de l'*explainable artificial intelligence* dont Henin [2021] fait partie qui asseoient leur explication sur un *jeu de données* capturant les entrées et les sorties de l'algorithme dans un scénario en « boîte noire », nous avons choisi de baser notre explication sur le programme et sa spécification directement, sans références à d'autres données, dans un scénario de « boîte blanche », qui est plus proche de la réalité des algorithmes administratifs en France. Bien évidemment, ces approches de l'explication ne sont pas mutuellement exclusives et il serait intéressant d'explorer leur combinaison, ce que nous laissons à des travaux futurs.

4.1 Expliquer par le droit, manuellement et automatiquement

La première étape de notre prototypage suppose de vérifier si la source primaire des règles de l'algorithme de calcul, c'est-à-dire les règles inscrites dans les textes de droit de la construction et de l'habitation, peut constituer un support d'explicabilité. Nous avons donc cherché à écrire à la main l'explication d'un calcul d'aides au logement, non pas dans le format succinct usité par les conseillères-gestionnaires des CAF et mentionné à la page 29, mais sous la forme d'une motivation juridique de décision administrative comme nous l'avons envisagée à la Section 2.2.

Méthodologie du cas pratique Pour élaborer notre explication, nous avons exploré la méthode du cas pratique, soit un exercice de mise en pratique des savoirs appris au cours d'une formation juridique classique. Le résultat du cas pratique se rapproche de celui d'une consultation juridique auprès d'un cabinet d'avocat et se caractérise par des développements littéraires parfois verbeux, mais rigoureux, permettant de tenir compte de la technicité de l'attribution et du calcul des droits qui nous intéresse ici. Nous aurions également envisagé le format d'écriture de la décision de justice. Cette forme littéraire présente également l'intérêt de lier le droit et les faits dans un format intelligible, mais elle est construite au regard des moyens apportés et des prestations des parties au procès instruit suivant le principe du contradictoire. Or, cela restreint le contenu de l'argumentation de la décision et ne permet pas de remplir pleinement les critères d'individualisation et de détail de l'explication.

Le cas pratique part d'un énoncé de faits décrivant une situation. L'exercice consiste à soulever une ou des questions de droit auxquelles une réponse, motivée par une argumentation normée, est suggérée. L'argumentation repose sur une suite de syllogismes dans laquelle s'appliquent des règles de droit. Le syllogisme se décompose en trois parties. Premièrement, la *majeure* expose le droit applicable à l'espèce et en définit succinctement les termes. Pour

chacune des règles applicables, la majeure en rappelle les conditions de mise en œuvre (notamment les modalités de preuve) et les exceptions, ainsi que les effets juridiques de la règle, et enfin la jurisprudence associée (évolution, absence de jurisprudence, position souple ou restrictive, etc.). La jurisprudence peut être complétée par des éléments de doctrine. Deuxièmement, la *mineure* du syllogisme applique le droit au cas d'espèce en montrant que les conditions développées dans la majeure ne sont pas réunies, ou au contraire sont bien remplies. Enfin, la *conclusion* répond à la question posée au début et tranche le cas d'espèce.

Le cas pratique nous a semblé constituer une forme pertinente d'explication individuelle d'une décision administrative algorithmique dans la mesure où il ressemble à une démonstration dont les éléments sont similaires aux éléments requis par l'article R. 311-1-3-2 du code des relations entre le public et l'administration mentionné dans la Section 2.2. Voici les correspondances auxquelles nous sommes arrivés :

1. « le degré et le mode de contribution du traitement algorithmique à la prise de décision » délimitent la partie de la décision prise en charge par l'algorithme et informent sur sa nature. Dans le cas pratique, c'est l'étendue de la question de droit posée au cas d'espèce et la méthode de réponse. Nous faisons donc correspondre ici l'étendue du cas pratique à ce qui est effectué par l'algorithme.
2. « les données traitées et leurs sources » correspondront aux faits du cas pratique.
3. « les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressée » correspondront aux mineures des syllogismes du cas pratique.
4. « les opérations effectuées par le traitement » correspondront aux majeures des syllogismes du cas pratique.

Notre cas pratique prendra pour acquises les déclarations de l'usagère sur des faits lors du remplissage d'un formulaire et ne cherchera pas à justifier la manière dont celle-ci a été rempli puisque cela ne relève pas du traitement algorithmique en tant que tel.

Une explication manuelle individualisée et détaillée, intelligible pour et par le droit L'annexe A page 57 contient le développement d'un exemple complet de cas pratique expliquant la décision automatisée de calcul de l'allocation logement pour une mère célibataire avec deux enfants à charge vivant à Saint-Pierre-et-Miquelon. Cette explication de huit pages constitue une explication individualisée, détaillée et intelligible. Elle est individualisée, car la méthodologie du cas pratique remet constamment en regard la règle générale (article D. 823-9, puis les articles D. 823-16 à 823-19 du code de la construction et de l'habitation, etc.) aux faits du cas d'espèce renseigné dans le formulaire grâce au syllogisme. Elle s'avère également détaillée, car le cas pratique constitue une démonstration de bout en bout qui ne néglige aucune étape du raisonnement et donc aucun élément qui influe (ou non) sur le résultat final. Enfin, la question de l'intelligibilité de cette démonstration est plus discutable dans la mesure où le cas pratique est écrit, durant 8 pages, dans le langage du droit. Cela emporte deux conséquences. Premièrement, ce sont les textes juridiques eux-mêmes qui définissent des calculs et des provisions complexes, et il faut composer avec leur rédaction qui se révèle souvent alambiquée ou opaque, y compris pour des juristes aguerries. Deuxièmement, la lecture d'un cas pratique nécessite des capacités analytiques normalement apprises au cours de la formation au droit. Cependant, justement parce que le cas pratique est un exercice méthodologique, sa visée pédagogique le rend accessible, à la condition que la lectrice fasse l'effort de se plonger dans les qualifications du droit et dispose de capacités de raisonnement déductives, dont le maniement du syllogisme. L'intérêt que nous retenons du cas pratique

réside toutefois dans sa familiarité pour les juristes, qui, lorsqu'elles sont avocates, membres d'associations spécialisées dans l'accès au droit, magistrates ou conseillères-gestionnaires dans les CAF, font partie, nous l'avons montré dans l'enquête de la Section 3, des usagères potentielles de technologies d'explicitation susceptibles de relayer à ou de vérifier pour le compte des allocataires concernées.

En choisissant le cas pratique, nous prenons une direction de prototypage très différente des travaux de designers évoqués page 19, car nous estimons que le droit est la source première de l'explication d'une décision administrative. Le caractère textuel, long et démonstratif de ce format d'explication n'a pas la force de séduction des travaux cités plus haut, mais il présente l'avantage de respecter les trois critères. Toutefois, nous n'excluons pas que ce format d'explication puisse être combiné à d'autres, par exemple dans une visée pédagogique sur l'accès au droit.

Maintenant que nous avons précisé avec le cas pratique un format d'explication souhaitable et que nous sommes capables d'en donner un modèle produit à la main par une juriste, il s'agit d'explorer des pistes de production *automatique* de telles explications. Qui dit production automatique dit système d'information et programme informatique ; aussi nous partirons pour nos prototypes à venir du programme Catala, conçu spécifiquement pour écrire des programmes informatiques à partir des règles de droit [Huttner et Merigoux 2022 ; Merigoux, Chataing et Protzenko 2021]. Grâce à notre maîtrise du langage de programmation et de sa chaîne de compilation, nous pouvons opportunément créer des systèmes de production d'explication valables pour *tous les programmes écrits dans ce langage* et qui ne sont donc pas spécifiques au cas des aides au logement dont nous continuerons à nous servir pour illustrer nos propos.

4.2 Expliquer par la traçabilité et le débogage

La deuxième piste de prototypage que nous avons suivie part directement du processus de développement des programmes informatiques, qu'ils soient écrits dans le langage Catala ou dans tout autre langage. Lors de ce développement, la programmeuse aux prises avec son implémentation de l'algorithme est fréquemment confronté à des bogues, c'est-à-dire des résultats de l'exécution du programme inattendus ou incorrects par rapport à sa spécification. La programmeuse peut alors adopter plusieurs stratégies. L'une d'elles consiste à « bidouiller » le programme aléatoirement ou en suivant son intuition jusqu'à ce que l'exécution du programme donne le résultat souhaité. Cependant, cette stratégie n'est pas très efficace et montre vite ses limites, car l'erreur dans le programme est souvent très décorrélée des indices contenus dans le résultat erroné de l'exécution. De manière générale, le débogage d'un programme est avant tout un problème de décalage entre le programme et la compréhension qu'en a la programmeuse. Une stratégie plus efficace consiste alors à formuler des hypothèses sur le programme et l'erreur, et à les tester grâce à des modifications bien choisies du programme, comme le relatait déjà Vessey [1985]. Chaque modification et exécution apporte plus d'informations sur le bogue et sa cause, de nature à guider la programmeuse vers la correction.

Durant ce processus ce débogage, il est essentiel d'explorer *ce que fait le programme* afin d'extraire des informations sur le bogue et sa cause. Aussi, une grande partie des modifications apportées au programme durant le processus de débogage consiste à faire afficher au programme plus d'informations sur les valeurs et les quantités intermédiaires qu'il manipule. L'affichage des valeurs intermédiaires pendant le débogage est tant indispensable aux

programmeuses que la plupart des langages de programmation offrent à leurs utilisatrices un arsenal d'outils qui automatisent l'affichage et l'exploration des valeurs intermédiaires du programme.

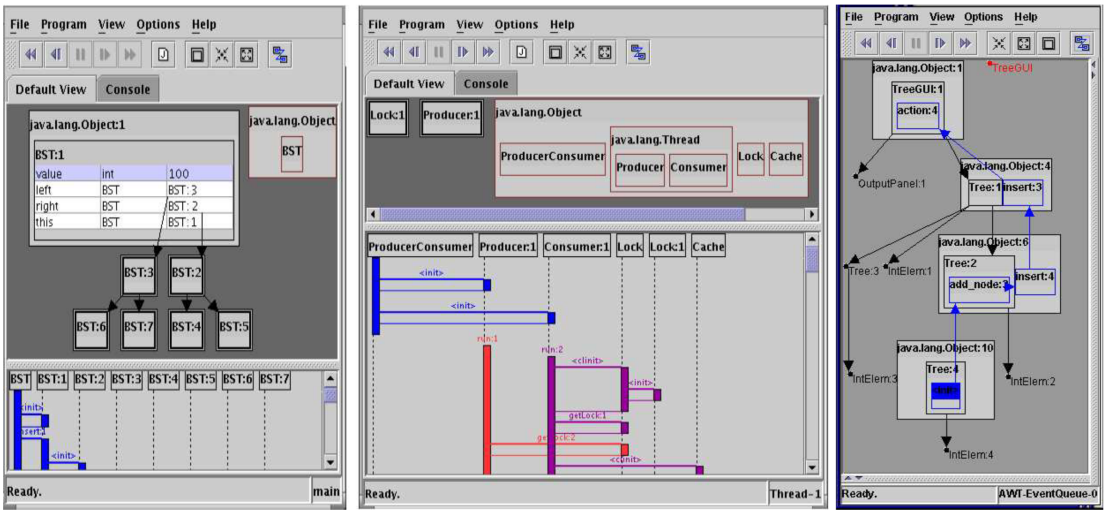


Figure 7 – Capture d'écran de JIVE extraite de [Gestwicki et Jayaraman 2004]. On y trouve les éléments de visualisation de l'exécution d'un programme Java avec la trace d'appel des méthodes des classes, ainsi que la représentation des objets créés par le programme et les références entre eux.

Ces outils peuvent opérer à différents niveaux : GDB [Stallman, Pesch, Shebs et al. 1988] par exemple permet d'examiner le comportement bas-niveau d'un programme, au sens des instructions assembleur exécutées par la machine, tandis que d'autres outils opèrent à plus haut niveau, permettant de visualiser l'exécution du programme en fonction des concepts du langage de programmation. Par exemple, l'outil JIVE de Gestwicki et Jayaraman [2004] permet de visualiser, pour des programmes Java sous forme graphique, les objets produits durant l'exécution d'un programme Java (voir Figure 7). Ainsi, il existe un grand nombre de manières de produire une visualisation de l'exécution du programme. Celles-ci peuvent être textuelles ou graphiques, statiques ou interactives, haut-niveau ou bas-niveau, orientées vers les interactions concurrentes des processus lancés par le programme, la consommation de mémoire ou les appels de fonction, etc. Cependant, elles partagent toutes l'objectif de faciliter la compréhension par une observatrice de ce que fait le programme.

Méthodologie de l'explication par la trace d'exécution Notre premier prototype d'explicabilité algorithmique pour Catala consiste donc en un outil de débogage basé sur la trace d'exécution similaire à ceux mentionnés ci-dessus. Dans un premier temps, ses utilisatrices principales seront les programmeuses qui écrivent le programme Catala. Catala étant un langage haut-niveau où le code source entremêle la spécification par le droit et les instructions informatiques, nous avons voulu produire une trace d'exécution qui entremêle également le droit et le code. Concrètement, lors de l'exécution du programme par l'interprète de Catala, un dispositif d'enregistrement (*log*) est appelé à des moments clés de l'exécution du programme : définition de la valeur d'une variable, appel à un champ d'application, à une fonction, etc. Il

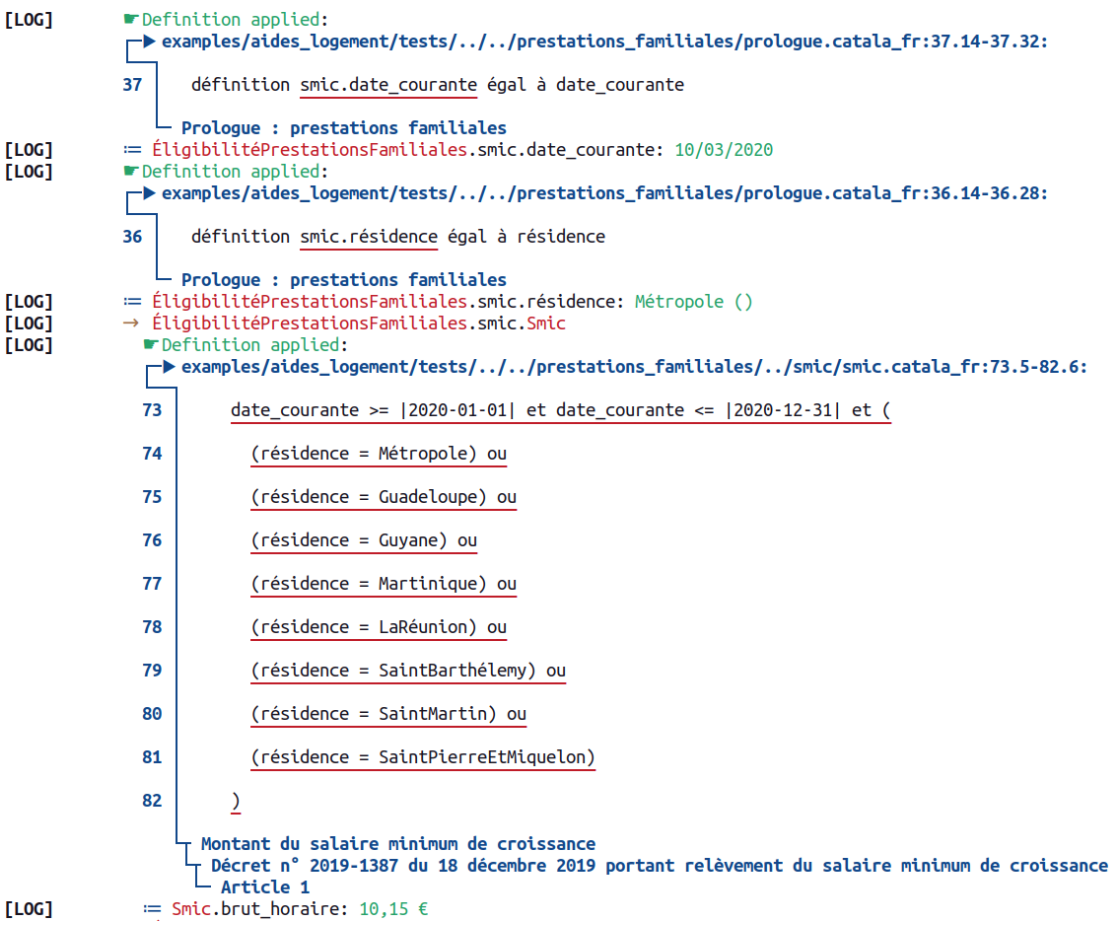


Figure 8 – Sortie textuelle de la trace d'exécution d'un programme Catala telle que produite par l'interpréteur avec l'option `--trace`. On y voit ici la séquence d'instructions permettant de déterminer la valeur du SMIC en fonction de la date courante et du lieu de résidence. En particulier, la valeur trouvée de 10,15 € brut de l'heure est définie par l'article 1 du décret n° 2019-1387 du 18 décembre 2019 portant relèvement du salaire minimum de croissance, et le code Catala se trouve à la ligne 73 du fichier `smic.catala_fr`.

est ainsi possible d'afficher chaque évènement enregistré de la sorte avec l'option `--trace` (voir Figure 8). À chaque définition de variable est associé le nom de la variable définie, sa valeur, et l'emplacement dans le code source correspondant à la définition précise choisie lors de l'exécution. Comme pour le cas pratique, cet emplacement permet de localiser la règle de droit précise utilisée pour valoriser la variable ; puisque le code source de la règle est situé à côté du texte de droit qui le spécifie, l'outil peut donc afficher dans l'exemple de la Figure 8 que le montant du SMIC au 1er janvier 2022 en métropole est déterminé par l'article 1 du décret n° 2021-1741 du 22 décembre 2021. Cet outil de débogage immédiatement accessible par les programmeuses pendant le développement s'est révélé utile pendant les séances de programmation en binôme informaticien-juriste et la conception des cas de tests en autonomie par des juristes (qui utilisent alors l'interface en ligne de commande du compilateur de Catala). L'avantage de cette explication algorithmique sous forme de trace d'exécution est sa production complètement automatique et à très faible coût puisque le mécanisme pour la concevoir est déjà présent dans le compilateur Catala. Ce mécanisme est d'ailleurs aisément transposable à d'autres langages pour peu que l'on puisse modifier les compilateurs. Cependant, l'austérité de l'interface textuelle de la ligne de commande n'est pas adaptée à un usage plus large de cette trace d'explication de l'exécution d'un programme Catala.

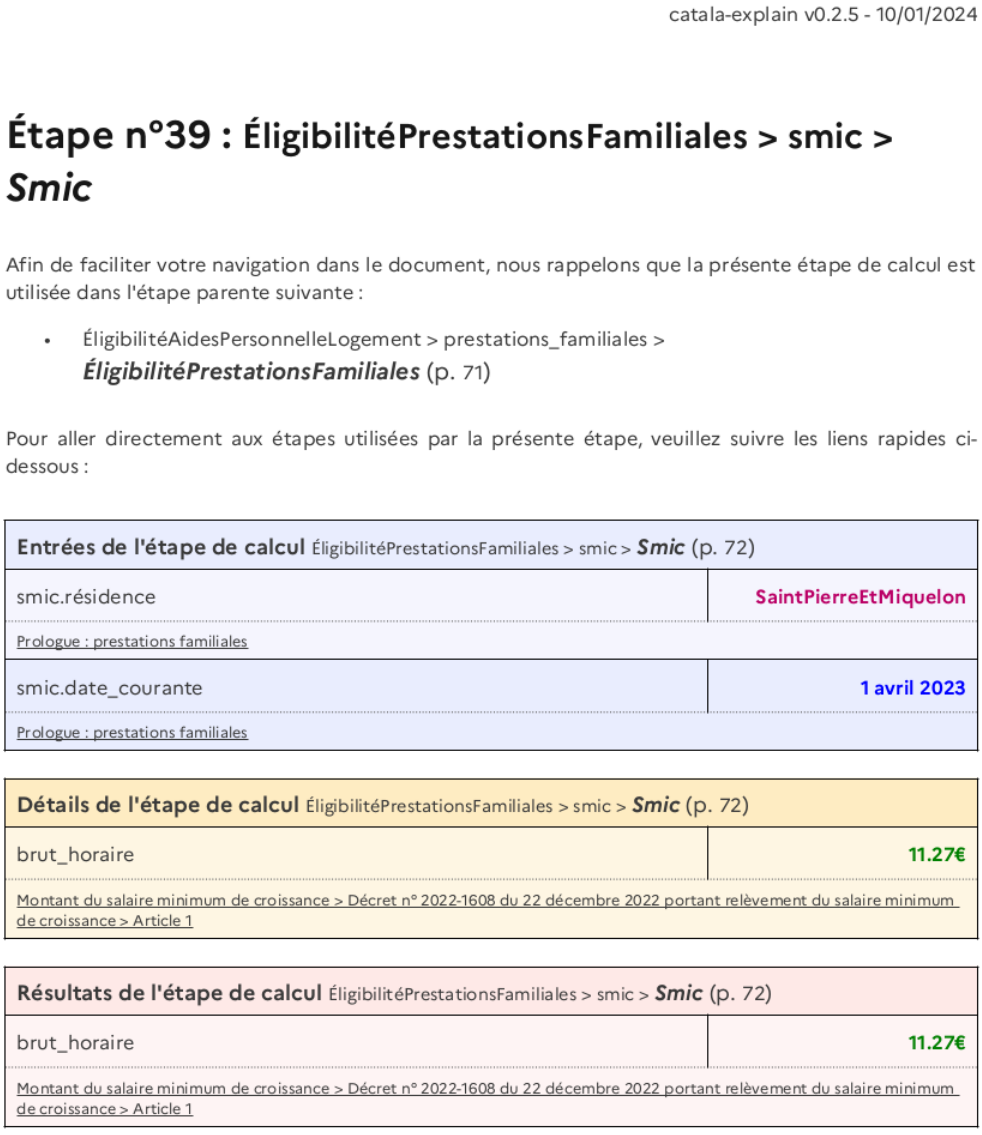
La trace d'exécution est un dispositif explicatif dont il nous a semblé pertinent de chercher à élargir le public au-delà des programmeuses et juristes directement impliquées dans le développement du programme Catala, en vertu de l'analyse des Sections 2 et 3 de ce rapport. Pour ce faire, nous avons cherché à afficher la trace d'exécution sous un format plus accessible par les non-programmeuses. Comme la trace d'exécution collectée lors de l'exécution d'un programme Catala se présente sous la forme d'un jeu de données qu'il est possible de manipuler, par exemple pour le distribuer au format JSON. Cette trace d'exécution au format JSON peut ensuite être consommée par plusieurs applications de visualisation qui l'afficheront avec différents styles graphiques et supports.

Nous avons produit pour l'instant deux applications de visualisation de la trace d'exécution. La première, illustrée Figure 9, centre sa représentation sur la structure du calcul en Catala. Une démonstration de cette visualisation est disponible à sur le site de Catala, et le code source correspondant est hébergé dans le dépôt `catala-website` accessible en licence libre sur GitHub. Le calcul en Catala est en effet découpé en plusieurs briques élémentaires appelées « champ d'application » qui correspondent à des notions plus ou moins indépendantes. Par exemple, un champ d'application permet de calculer la valeur du SMIC en fonction de la date et du lieu de résidence, un autre permet de déterminer l'éligibilité à l'allocation logement en fonction des paramètres du ménage, etc. Ces champs d'application peuvent dépendre les uns des autres, puisque par exemple la prise en compte d'une enfant pour l'éligibilité peut dépendre de si sa rémunération va dépasser une certaine somme qui dépend de la valeur du SMIC. Ainsi la première visualisation prend la forme d'une page Web qui présente les événements de la trace en les imbriquant dans les champs d'application associés. Un champ d'application va par exemple contenir toutes les définitions de ses variables, qui peuvent elles-mêmes contenir les appels d'autres champs d'application nécessaire au calcul de leur valeur, etc. L'imbrication permet de concevoir une interface où tout est replié à l'origine, et où l'utilisatrice peut cliquer pour dévoiler une étape du calcul en particulier, afin de naviguer vers l'événement précis de la trace qui l'intéresse. Déjà plus exploitable pour les juristes que l'interface en ligne de commande, cette visualisation présente l'inconvénient de rapidement devenir un « cliquodrôme » et d'encombrer la visualisation par un emboîtement trop complexe des écrans correspondant aux événements imbriqués du calcul.

Figure 9 – Extrait de l'interface Web de visualisation de la trace d'exécution d'un programme Catala, accessible sur le site de Catala (dont le code source est disponible dans le dépôt `catala-website`).

La deuxième interface de visualisation, illustrée Figure 10, est conçue directement en contraste avec la première, et revient aux fondamentaux des documents d'explicabilité présentés dans la Section 2.2. Une démonstration de cette deuxième interface est disponible à l'adresse `https://code.gouv.fr/demos/catala/`, son code source se trouvant dans le dépôt `catala-dsfr` accessible en licence libre sur GitHub. Elle a bénéficié d'un financement de la DINUM pour son développement. Les emboîtements d'événements et de champ d'application, plutôt que d'être présentés imbriqués, sont ici linéarisés en étapes de calcul qui font références les unes aux autres. Les étapes de calcul sont ainsi présentées dans un document de traitement de texte comportant des tableaux, à la manière des tableaux des fiches de paie et des avis d'imposition. Chaque étape de calcul se voit attribuer un numéro et le document permet à l'autrice de naviguer entre les étapes de calcul en suivant leurs numéros ou le numéro de page auxquelles elles sont présentées. Le format traitement de texte de ce document assure une diffusion potentiellement large, y compris par une exportation au format PDF, ce format de document s'insérant sans problème dans n'importe quel processus de communication administrative (demande d'accès au dossier par exemple).

Une explication automatique individualisée et détaillée, mais dont l'intelligibilité reste à travailler La production automatique d'explications individualisées et détaillées à partir de la trace d'exécution du programme est donc techniquement possible, grâce à la mutuali-



catala-explain v0.2.5 - 10/01/2024

Étape n°39 : ÉligibilitéPrestationsFamiliales > smic > *Smic*

Afin de faciliter votre navigation dans le document, nous rappelons que la présente étape de calcul est utilisée dans l'étape parente suivante :

- ÉligibilitéAidesPersonnelleLogement > prestations_familiales > ***ÉligibilitéPrestationsFamiliales*** (p. 71)

Pour aller directement aux étapes utilisées par la présente étape, veuillez suivre les liens rapides ci-dessous :

Entrées de l'étape de calcul ÉligibilitéPrestationsFamiliales > smic > ***Smic*** (p. 72)	
smic.résidence	SaintPierreEtMiquelon
Prologue : prestations familiales	
smic.date_courante	1 avril 2023
Prologue : prestations familiales	

Détails de l'étape de calcul ÉligibilitéPrestationsFamiliales > smic > ***Smic*** (p. 72)	
brut_horaire	11.27€
Montant du salaire minimum de croissance > Décret n° 2022-1608 du 22 décembre 2022 portant relèvement du salaire minimum de croissance > Article 1	

Résultats de l'étape de calcul ÉligibilitéPrestationsFamiliales > smic > ***Smic*** (p. 72)	
brut_horaire	11.27€
Montant du salaire minimum de croissance > Décret n° 2022-1608 du 22 décembre 2022 portant relèvement du salaire minimum de croissance > Article 1	

Figure 10 – Extrait de l'interface en document textuel de la visualisation de la trace d'exécution d'un programme Catala, produite par `catala-dsfr` qui utilise lui-même la librairie `catala-explain` pour générer les documents texte.

sation de l'infrastructure technique entre le calcul et son explication. Ce principe est déjà à l'œuvre pour la production des avis d'imposition comme celui de la page 17. Dans notre prototype constitué par ces deux interfaces de visualisation de la trace d'exécution, la valeur ajoutée de Catala est de relier entre eux événements de l'exécution du programme et dispositions législatives ou réglementaires qui le spécifient. Ce lien nous permet de boucler la boucle entre le fondement de la décision automatisée et son explication (qui permet éventuellement de la contester), à travers le dispositif technique nécessaire au calcul massif de ces décisions à l'échelle de la population. Ce lien permet aussi d'aligner le contrôle interne et externe du système d'information, puisque d'éventuelles erreurs dans la prise de décision peuvent être retracées au moyen d'un processus classique de débogage à un endroit précis de l'explication, qui correspond donc à un emplacement du code la disposition législative ou réglementaire qu'il est censé suivre.

Cependant, l'intelligibilité de l'explication produite à partir de la trace d'exécution laisse à désirer. Le design de ce prototype, constitué par ces deux interfaces de visualisation de la trace d'exécution, n'est pas finalisé. En l'état, la trace d'exécution reste un objet informatique éminemment technique et s'en servir pour expliquer l'algorithme à une allocataire concernée par le calcul nécessite un travail de traduction des concepts et symboles informatiques vers des concepts et symboles que cette dernière serait susceptible de comprendre. Autrement dit, l'explication par la trace est individualisée et détaillée par défaut, mais pas intelligible en dehors du cercle réduit d'une direction des systèmes d'information. En ce sens, elle serait d'abord utile pour certaines activités de contrôle interne à la DSI de la CNAF, comme le débogage ou la mise à jour de la législation. Ce travail de traduction de concepts et symboles dans l'affichage de la visualisation est d'ailleurs très difficile à faire *in abstracto*, c'est-à-dire indépendamment de l'algorithme que l'on cherche à expliquer. Concrètement, nous avons évité dans notre prototypage d'incorporer dans nos visualisations des éléments spécifiques aux aides au logement, nous reposant plutôt sur des concepts et symboles génériques à tous les programmes Catala (« étape de calcul », « variable », etc.).

Un dispositif d'explicabilité algorithmique respectueux des exigences légales et à destination des personnes concernées par la décision se devra donc d'être personnalisé à l'algorithme que l'on cherche à expliquer, car c'est le seul moyen de tisser un lien fin et précis entre les notions informatiques produites automatiquement et les objets sujets de l'algorithme. Par exemple, notre trace quand elle calcule l'éligibilité aux aides au logement des enfants à charge n'est pas capable de donner un prénom à ces enfants ni de demander à l'utilisatrice de donner un prénom à chacun des enfants en entrée afin que la visualisation puisse reprendre ce prénom dans ses explications. Aussi, en concevant un outil générique, nous avons voulu laisser des opportunités facilement actionnables de personnalisation de l'explication à un algorithme particulier. Une première opportunité vient de l'architecture même du dispositif, qui manipule la trace au format JSON de manière symbolique et qui permet donc d'ajouter des règles personnalisées de traitement de cette trace. Mais nous avons aussi envisagé de modifier nos interfaces de visualisation pour que ces applications prennent en entrée, non seulement la trace d'exécution, mais aussi la liste des règles de présentation de cette trace personnalisée par rapport à l'algorithme donné, ce qui fournirait ainsi un cadriciel intégré de production automatique de ces visualisations. Cette dernière fonctionnalité fera l'objet de futurs développements.

Une autre limitation importante des explications générées automatiquement par nos interfaces de visualisation de la trace d'exécution tient en leur longueur. Par exemple, le PDF expliquant le calcul des aides au logement dont la Figure 10 est un extrait compte pas moins de

210 pages. Puisque chaque petite étape du calcul est détaillée à l'extrême, l'explication risque de noyer le lecteur sous une montagne de détails techniques sur les plis du calcul, qui rend difficile l'extraction de l'information précise dont il aurait besoin. Des stratégies pourraient être employées pour guider la lecture, comme la factorisation d'étapes de calcul identiques dans l'explication. Cependant, cette limitation nous a paru suffisamment sérieuse pour que nous envisagions une alternative technique plus radicale pour la production d'explications automatiques de programmes écrits en Catala.

4.3 Expliquer par l'évaluation paresseuse du calcul

La longueur des explications générées par les outils précédents tient à des spécificités du calcul qui n'influencent pas forcément directement le résultat final. Par exemple, dans le calcul des aides au logement pour les secteurs logement-foyer et accession à la propriété, le programme Catala de Merigoux et Slimani [2022] renvoie systématiquement le montant le plus élevé entre l'allocation logement et l'aide personnalisée au logement quand l'utilisatrice est éligible aux deux (comportement discuté dans la Section 3.3 de Merigoux, Alauzen et Slimani [2023]). Mais pour expliquer le montant final de l'aide renvoyé à l'utilisatrice, il n'est peut-être pas souhaitable d'expliquer la manière dont l'allocation logement est calculée si l'aide personnalisée au logement, plus élevée, est l'aide effectivement versée. Autrement dit, le programme Catala explore pour les besoins du calcul des branches qui seront par la suite abandonnées. Ainsi, une explication individualisée peut rester détaillée, mais devenir plus intelligible si l'on retire de l'explication toutes ces branches abandonnées. De plus, la structure du calcul avec les champs d'applications qui s'emboîtent les uns dans les autres n'est pas directement utile à l'explication.

Une explication plus concise consisterait ainsi en une simple suite d'étapes de calcul qui mèneraient directement au résultat final, effaçant tous les embranchements et les hésitations superflues que l'on peut considérer comme autant d'artefacts de calcul qui n'auraient pas eu leur place dans un cas pratique écrit à la main à la manière de celui de l'Annexe A décrit en Section 4.1. Alors, comment détecter automatiquement ces branches abandonnées et linéariser ainsi le calcul ? La trace de calcul exploitée en Section 4.2 ne contient pas assez d'information pour produire une explication de cette sorte. En effet, elle contient les événements successifs du calcul, mais pas les raisons pour lesquelles ces événements sont survenus plutôt que d'autres.

Il faut donc revenir à la source du calcul et donc à l'exécution du programme lui-même pour en extraire davantage d'information. On pourrait continuer à enrichir la trace d'exécution du programme en rajoutant pour chaque événement du calcul la liste des raisons qui ont conduit à cet événement, la formule de calcul associée à l'événement, etc. À partir de cette trace enrichie, on pourrait ensuite effectuer un traitement complexe qui permettrait de déterminer si telle ou telle partie de la trace est pertinente ou pas, et de la supprimer de l'explication le cas échéant.

Cependant, il existe une manière alternative et plus élégante de conceptualiser et d'implémenter cette opération de nettoyage des branches abandonnées du calcul. Plutôt que de modifier et traiter la trace d'exécution, il s'agit de modifier directement *la manière dont le programme s'exécute*.

Méthodologie de l'explication par évaluation paresseuse En effet, l'exécution d'un programme Catala est définie formellement au niveau du calcul par défaut par une sémantique opé-

rationnelle présentée dans Merigoux, Chataing et Protzenko [2021] [4]. La sémantique opérationnelle est un ensemble de règles précises qui définissent, dans toutes les situations, comment l'on doit en pratique effectuer le calcul défini par un programme Catala. Sous cette sémantique par exemple, une règle nous dit que le programme `3+2` s'évalue en `5`, qui est bien le résultat de l'addition de 3 et 2. De même, une autre règle nous dit si l'on considère le programme `soit x égal à 2 dans 1 + x`, alors il faut remplacer `x` par `2` pour évaluer le résultat `1 + 2` qui s'évalue lui-même vers `3` en vertu de la règle précédente.

Mais, en changeant de sémantique d'évaluation, on modifie les règles par lesquelles on calcule le résultat du programme. Pour répliquer la production d'explications avec l'abandon des branches inutiles, nous avons défini une nouvelle sémantique d'évaluation [5] de Catala dans laquelle le résultat du programme `3+2` est... `3+2`. En effet, puisqu'il s'agit d'expliquer le calcul, il ne faut pas calculer les opérations mathématiques, mais plutôt les laisser telles quelles pour expliquer d'où le résultat provient. Parce qu'elle n'évalue pas complètement les calculs, cette sémantique d'évaluation est dite *paresseuse*, par opposition aux sémantiques *enthousiastes*. Par contre, cette nouvelle sémantique élimine bien les branches abandonnées de l'évaluation pour rendre l'explication plus concise. Une règle prévoit en effet que toutes les valeurs nécessaires à faire un choix dans le programme (prendre une allocation plutôt qu'une autre, plafonner ou pas, etc.) doivent elles être calculées complètement, pour qu'on puisse sélectionner le résultat du choix associé et abandonner les options du choix non-sélectionnées. Concrètement, considérons le programme

```
si âge > 18 alors 400 € + 50 € sinon 0 €
```

On suppose ici que `âge = 20`. Comme notre programme va devoir faire le choix entre `400 € + 50 €` et `0 €`, la règle ci-dessus nous dit qu'il faut évaluer si `âge > 18`; c'est le cas, donc on va pouvoir évaluer directement ce programme vers le résultat du choix, soit `400 € + 50 €`. Dans l'explication, l'utilisatrice ne verra pas donc trace de la potentialité du `0 €` qui est évitée puisque son `âge` est supérieur à `18`.

La Figure 11 illustre le résultat de cette évaluation paresseuse. Plutôt que de visualiser le résultat de l'évaluation paresseuse sous la forme textuelle de son résultat, nous avons préféré l'afficher sous la forme d'un graphe, qui matérialise les relations de dépendance entre les variables du programme. En haut, le début du calcul avec les entrées de l'utilisatrice en orange, et le programme progresse en allant vers le bas. On peut remarquer que la réduction de la taille de l'explication est spectaculaire, puisque les 210 pages de la Section 4.2 tiennent maintenant sur une page (certes, un peu condensée !). Cependant, la stratégie d'élimination des branches abandonnées est allée un peu trop loin dans la mesure où elle a retiré des éléments cruciaux pour comprendre le calcul : comme sur la fiche de paie de la Section 2.2, on a une suite de calculs aboutissant au bon résultat, mais on ne peut pas comprendre pourquoi ni comment les valeurs dans ces calculs ont été choisies. Concrètement, on remarque à la fin du calcul une étrange modulation de la valeur de l'aide par un facteur proportionnel à `2023−2026`. En fait, cette étape du calcul est présente parce que, dans l'exemple, le ménage réside à Saint-Pierre-et-Miquelon. Mais comme le test de présence à Saint-Pierre-et-Miquelon ne figurait que dans la condition qui active cette étape de calcul, il a été éliminé de l'explication.

4. Cette sémantique suit le modèle de l'évaluation enthousiaste avec appel par valeur du lambda-calcul, augmentée des règles relatives au terme par défaut

5. L'Annexe B décrit formellement l'ensemble des règles de cette nouvelle sémantique d'évaluation.

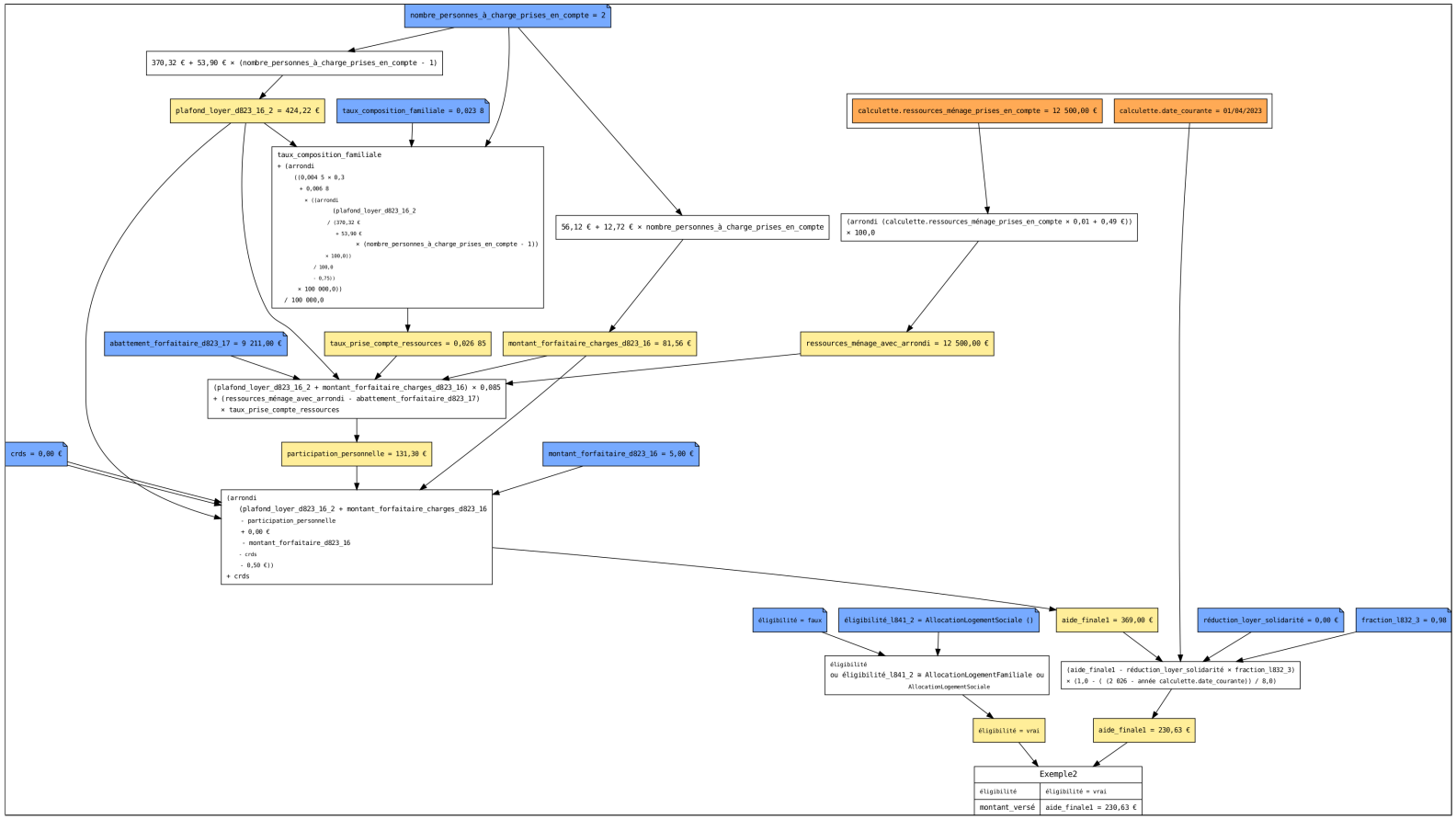


Figure 11 – Explication du calcul des aides au logement par évaluation partielle et paresseuse, visualisée sous forme de graphe de dépendance des variables.

Cette expérience permet de relativiser l'objectif de concision de l'explication. Comme nous l'avons vu précédemment, l'intelligibilité d'une explication est en tension avec son caractère détaillé. D'où la nécessité réitérée ici de produire plusieurs niveaux d'explication, qui rentrent plus ou moins dans le détail afin de guider l'utilisatrice, qu'elle soit avocate ou programmeuse, vers la compréhension du point précis qui l'intéresse. Ici, notre contribution est une méthode de production *automatique* et *indépendante du programme* (tant qu'il est écrit en Catala) d'explications plus ou moins détaillées de son exécution. En effet, l'évaluation paresseuse peut être légèrement modifiée afin de ne pas éliminer totalement les choix et ainsi de faire apparaître les raisons qui président à l'établissement de telle ou telle valeur dans l'évaluation, comme c'est illustré dans la Figure 12.

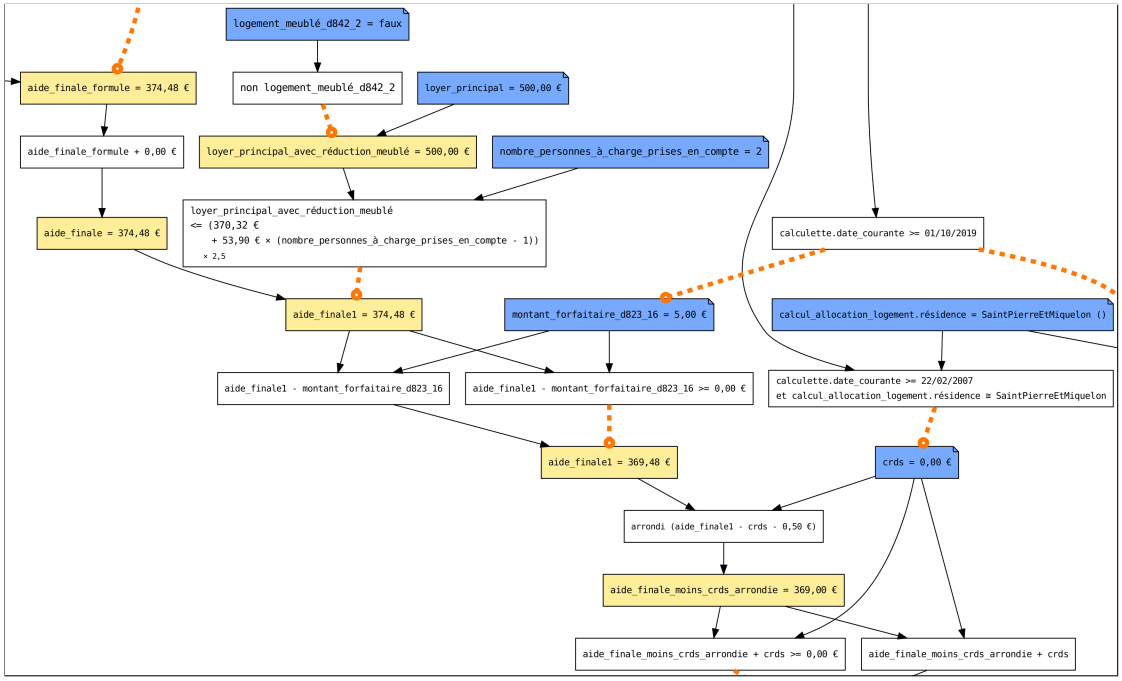


Figure 12 – Extrait d'un détail de la visualisation de la Figure 11 augmentée des relations logiques entre les conditions et les variables en trait pointillé orange.

Une explication automatique individualisée, détaillée et intelligible ? Actuellement, les explications générées par l'évaluation paresseuse des programmes Catala manquent d'une visualisation tournée vers les utilisatrices comme celle de la Section 4.2 relative à la trace d'exécution ; le graphe est en effet difficilement lisible et compréhensible en dehors du monde restreint des informaticiennes spécialistes de Catala. Cependant, il serait possible de réutiliser le travail de visualisation précédemment effectué en générant, par l'évaluation paresseuse, des traces d'exécution au même format que celles produites par l'évaluation classique des programmes Catala.

L'évaluation paresseuse ne résout pas le problème évoqué également dans la Section 4.2 de personnalisation de l'explication en fonction de l'algorithme expliqué. La technique ici présentée est encore une fois générique et ne dépend pas du programme Catala exécuté. Enfin,

un désavantage actuel de cette technique de production d'évaluation est qu'elle nécessite l'interprétation du programme Catala à l'intérieur du compilateur, alors que la trace d'exécution de la Section 4.2 pouvait être générée à l'intérieur de n'importe quel langage cible de compilation de Catala. Par exemple, les traces traitées par les visualisations de la Section 4.2 sont générées par la version compilée en JavaScript du programme Catala originel. L'utilisation de l'interpréteur par rapport à une version compilée du code entraîne une chute drastique de la performance ainsi que la nécessité d'embarquer dans l'application de génération d'explications une copie du compilateur Catala et de ses systèmes auxiliaires (implémentés en OCaml). Nous laissons à des travaux futurs l'intégration de l'évaluation paresseuse à la version compilée des programmes Catala pour passer outre les problèmes de performance et ne pas nous encombrer de l'interpréteur.

Reprenons cette série de prototypes en nous demandant la nature de leur contribution à la discussion sur l'explicabilité des algorithmes publics. Nous avons d'abord produit manuellement le format d'explications pour et par le droit qui correspond aux manières de faire des professionnelles du droit et qui peut être lisible par les intermédiaires du droit identifiées dans la Section 3. Par analogie, nous avons relié le format du cas pratique juridique à celui des traces d'exécution des programmes informatiques, qui sont des explications que l'on peut produire automatiquement et que nous avons exploitées sous diverses interfaces dans notre premier prototype technique avec Catala. Constatant la longueur des explications générées et le manque de lien entre les concepts informatiques de la trace et les concepts manipulés par le droit, nous avons suggéré deux directions d'amélioration du prototype initial. La première consiste à personnaliser la présentation de la trace d'exécution selon des règles qui dépendent de l'algorithme expliqué, et la seconde à masquer des branches abandonnées du calcul qui n'influent pas sur le résultat final. Nous avons ensuite démontré la faisabilité de la seconde amélioration avec un deuxième prototype. Aussi, le travail présenté dans ce rapport démontre qu'il est techniquement faisable, sur la base de programmes écrits d'une manière liant le droit et le code comme le langage Catala, de générer automatiquement des explications individualisées, détaillées et possédant une certaine intelligibilité sur la base de droit que suit l'algorithme. La transformation de ces prototypes conçus pour alimenter le débat sur l'explicabilité des algorithmes publics en des outils industriels dont pourraient se saisir les administrations pourraient constituer un prolongement du rapport de recherche.

5 Conclusion

Dans ce rapport, nous avons présenté le résultat d'une recherche collective et interdisciplinaire partant du droit, supposant d'aller vers la sociologie et informatique pour proposer des pistes de recherches opérationnelles sur l'explicabilité des algorithmes publics. Partant de l'idée selon laquelle la publication en source ouverte est une condition nécessaire, mais non suffisante de la transparence de l'administration vis-à-vis du public, nous avons construit un état des lieux des formes d'explication des algorithmes publics disponibles. Nous avons ainsi distingué finement les formes d'explication du calcul et leurs effets en les rapportant à la triple exigence établie par le droit de l'informatique et des libertés. Nous avons ensuite enquêté sur un cas concret, l'algorithme de calcul des aides au logement, qui fait l'objet de controverses. Cette exploration empirique nous a permis de clarifier les usages et les usagères de l'explication des décisions administratives, à la fois au cours du processus de prise de décision et de contestation de la décision. Le détour par l'enquête nous a permis de concevoir trois types d'explication d'un algorithme public, un prototype manuel reposant sur la méthode du cas pratique juridique et deux prototypes plus fidèles à l'exécution technique

de l'algorithme. Ces prototypes démontrent la faisabilité technique de la production d'explications individualisées, détaillées et intelligibles par le droit ; et nous réaffirmons la thèse selon laquelle la mise à disposition de telles explications garantirait un meilleur contrôle interne et externe des décisions automatiques prises par une administration donnée.

Plutôt que d'espérer expliquer les algorithmes à n'importe qui, nous avons déterminé qu'il était plus opportun de viser d'abord les personnes directement impliquées dans la conception du programme et sa contestation. Sa conclusion quelque peu pessimiste quant à l'utilité de la redevabilité algorithmique ne doit cependant pas obérer notre volonté d'aller au bout de l'application des dispositifs de la loi informatiques et libertés et de ramener la décision automatisée dans le cadre qui régule l'action de l'administration pour en garantir sa justesse et en augmenter la lisibilité. Les prototypes que nous avons conçus dans le cadre de ce travail ne sont pas techniquement révolutionnaires, et reposent tous sur des concepts bien établis en informatique et étayés par des travaux antérieurs. On pourrait même qualifier d'ingénierie avancée la plupart de nos développements, sans pour autant en anéantir la portée. Le développement et le maintien de ces cœurs de systèmes d'information responsables du calcul et de l'explication des décisions automatisées spécifiées tout en partie par le droit restent un défi majeur auquel doivent faire face les administrations au quotidien, et l'objet technique qu'est Catala doit engranger beaucoup d'ingénierie pour sortir peu à peu de l'enceinte du laboratoire et devenir un potentiel candidat au déploiement en production dans ces systèmes décidément critiques pour notre société.

Remerciements

Nous remercions toutes les personnes rencontrées dans le cadre de cette enquête d'avoir pris le temps de répondre à nos questions, souvent techniques, de nous avoir transmis de la documentation et d'avoir relu des versions intermédiaires du présent rapport. Elles sont évidemment déchargées de toute de responsabilité ; nos propos et les éventuelles erreurs nous engagent seuls. Caroline Lequesne-Roth nous a facilité l'accès au terrain et Liane Huttner nous a permis de nous adosser sur les apports de sa thèse sur les algorithmes décisionnels ; nous les remercions toutes deux chaleureusement de leur collaboration. Nous avons présenté les premiers résultats de cette recherche, le 9 juin 2023, lors du séminaire Blue Hats organisé par Etalab ; un grand merci aux organisatrices et aux participantes d'avoir nourri de leurs questions ce travail au long cours.

Financement

Dans le cadre des travaux menés sur la transparence des algorithmes, la direction interministérielle du numérique (DINUM) a financé au cours de l'année 2023 la conception du prototype d'explicabilité présenté dans la Section 4.2 par un bon de commande de 36 465 euros HT auprès de la société détentrice du marché public, Malt. L'intégralité du résultat de cette prestation est disponible sur le site de la mission logiciels libres de la DINUM à l'adresse suivante : https://code.gouv.fr/demos/catala [6].

6. Le code source est disponible sur le dépôt GitHub https://github.com/CatalaLang/catala-dsfr

Références

Littérature scientifique

Abel, Richard L. (1998). « Speaking Law to Power. Occasions for Cause Lawyering ». In : *Cause Lawyering : Political Commitments and Professional Responsibilities*. Sous la dir. d'Austin Sarat et Stuart A. Scheingold. Oxford University Press, p. 69-117.

Bellotti, Marianne (2021). *Kill It with Fire : Manage Aging Computer Systems (and Future Proof Modern Ones)*. San Francisco : No Starch Press.

Bench-Capon, Trevor JM et Frans P Coenen (1992). « Isomorphism and legal knowledge based systems ». In : *Artificial Intelligence and Law* 1.1, p. 65-86.

Berners-Lee, Tim, James Hendler et Ora Lassila (2001). « The semantic web ». In : *Scientific american* 284.5, p. 34-43.

Bertrand, Astrid et al. (2022). « How cognitive biases affect XAI-assisted decision-making : A systematic review ». In : *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, p. 78-91.

Bidet, Alexandra (2010). « Qu'est-ce que le vrai boulot ? Le cas d'un groupe de techniciens ». In : *Sociétés contemporaines* 78.02, p. 115-135.

Brooks, Frederick P (1974). « The mythical man-month ». In : *Datamation* 20.12, p. 44-52.

Burrell, Jenna (2016). « How the machine 'thinks' : Understanding opacity in machine learning algorithms ». In : *Big Data & Society* 3.1, p. 2053951715622512. doi : 10.1177/2053951715622512. eprint : https://doi.org/10.1177/2053951715622512. url : https://doi.org/10.1177/2053951715622512.

Cardon, Dominique, Jean-Philippe Cointet et Antoine Mazières (2018). « La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle ». FR. In : *Réseaux* 211.5, p. 173-220. doi : 10.3917/res.211.0173. url : https://www.cairn.info/revue-reseaux-2018-5-page-173.htm.

Cath, Corinne et Fieke Jansen (2021). *Dutch Comfort : The limits of AI governance through municipal registers*. arXiv : 2109.02944 [cs.AI].

Cellard, Loup (2020). « Theatres of algorithmic transparency : a post-digital ethnography ». Thèse de doct. University of Warwick.

Chabbat, Bertrand (1997). « Modélisation multiparadigme de textes réglementaires ». Thèse de doctorat dirigée par Pinon, Jean-Marie et Ou-Halima, Mohamed Informatique Lyon, INSA 1997. Thèse de doct. INSA Lyon, 1 vol. (391 p.) url : http://www.theses.fr/1997ISAL0118.

Denis, Jérôme et David Pontille (2012). « Travailleurs de l'écrit, matières de l'information ». In : *Revue d'anthropologie des connaissances* 6.6-1.

Deville, Clara (2023). *L'Etat social à distance - Dématérialisation et accès aux droits des classes populaires rurales*. Sous la dir. d'Action Publique. Éditions du Croquant.

Escher, Nel et Nikola Banovic (mai 2020). « Exposing Error in Poverty Management Technology : A Method for Auditing Government Benefits Screening Tools ». In : *Proc. ACM Hum.-Comput. Interact.* 4.CSCW1. doi : 10.1145/3392874. url : https://doi.org/10.1145/3392874.

Ewick, Patricia et Susan S Silbey (1998). *The common place of law : Stories from everyday life*. University of Chicago Press.

Gélédan, Fabien (juill. 2021). « Le design peut-il ré-enchanter l'action publique ? » Version longue d'un texte publié dans Action publique. Recherche et pratiques, n°7 Printemps 2020. url : https://hal.science/hal-03289398.

Gestwicki, Paul V et Bharat Jayaraman (2004). « Jive : Java interactive visualization environment ». In : *Companion to the 19th annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications*, p. 226-228.

Goëta, Samuel et Tim Davies (déc. 2016). « The Daily Shaping of State Transparency : Standards, Machine-Readability and the Configuration of Open Government Data Policies ». In : *Science & Technology Studies* 29.4, p. 10-30. doi : 10.23987/sts.60221. url : https://sciencetechnologystudies.journal.fi/article/view/60221.

Grimmelmann, James (mai 2023). « The Structure and Legal Interpretation of Computer Programs ». In : *Journal of Cross-disciplinary Research in Computational Law* 1.3. url : https://journalcrcl.org/crcl/article/view/19.

Henin, Clément (2021). « Expliquer et justifier les systèmes de décisions algorithmiques ». Thèse de doctorat dirigée par Castelluccia, Claude et Le Métayer, Daniel Informatique Lyon 2021. Thèse de doct. url : http://www.theses.fr/2021LYSEI058.

Henin, Clément et Daniel Le Métayer (nov. 2021). « A framework to contest and justify algorithmic decisions ». In : *AI and Ethics* 1.4, p. 463-476. issn : 2730-5961. doi : 10.1007/s43681-021-00054-3. url : https://doi.org/10.1007/s43681-021-00054-3.

– (déc. 2022). « Beyond explainability : justifiability and contestability of algorithmic decision systems ». In : *AI & SOCIETY* 37.4, p. 1397-1410. issn : 1435-5655. doi : 10.1007/s00146-021-01251-8. url : https://doi.org/10.1007/s00146-021-01251-8.

Hildebrandt, Mireille (nov. 2020). « Code-driven Law : Freezing the Future and Scaling the Past ». en. In : *Is Law Computable ? : Critical Perspectives on Law and Artificial Intelligence*. Sous la dir. de Simon Deakin et Christopher Markou. Oxford : Hart Publishing, p. 67-84. isbn : 978-1-50993-709-7.

Hitzler, Pascal (2021). « A review of the semantic web field ». In : *Communications of the ACM* 64.2, p. 76-83.

Huttner, Liane (2022). « La décision de l'algorithme. Étude de droit privé sur les relations entre l'humain et la machine. » Thèse de doct. Université Panthéon-Sorbonne.

Huttner, Liane et Denis Merigoux (août 2022). « Catala : Moving Towards the Future of Legal Expert Systems ». In : *Artificial Intelligence and Law*. doi : 10.1007/s10506-022-09328-5. url : https://hal.inria.fr/hal-02936606.

Jouve, David (2003). « Modélisation sémantique de la réglementation ». Thèse de doctorat dirigée par Pinon, Jean-Marie et Amghar, Youssef Informatique Lyon, INSA 2003. Thèse de doct. INSA Lyon, XIV-257 p. url : http://www.theses.fr/2003ISAL0071.

Kennan, Ariel et Sara Soka (2022). *Benefit Eligibility Rules as Code : Reducing the Gap Between Policy and Service Delivery for the Safety Net*. Rapp. tech. Beeck Center, Georgetown University. url : https://beeckcenter.georgetown.edu/report/benefit-eligibility-rules-as-code/.

Kounowski, Gilles (2002). « L'informatique et le système d'information des Allocations familiales ». In : *Revue des politiques sociales et familiales* 68.1. Included in a thematic issue : La branche Famille de la Sécurité sociale. Rétrospectives et prospectives, p. 49-72. doi : 10.3406/caf.2002.1018. url : https://www.persee.fr/doc/caf_1149-1590_2002_num_68_1_1018.

Krajewski, Markus (2023). « Source Code Criticism : On Programming as a Cultural Technique and its Judicial Linkages ». In : *Journal of Cross-disciplinary Research in Computational Law* 1.3. (forthcoming).

Kroll, Joshua A et al. (2017). « Accountable algorithms ». In : *University of Pennsylvania Law Review* 165, p. 633.

Lombrozo, Tania (2006). « The structure and function of explanations ». In : *Trends in cognitive sciences* 10.10, p. 464-470.

Merigoux, Denis (sept. 2022). *Observations sur le calcul des aides au logement*. Research Report RR-9485. Inria Paris, p. 27. url : https://hal.inria.fr/hal-03781578.

– (jan. 2023). « Experience report : implementing a real-world, medium-sized program derived from a legislative specification ». In : *Programming Languages and the Law 2023 (affiliated with POPL)*. Boston (MA), United States. url : https://inria.hal.science/hal-03933574.

Merigoux, Denis, Marie Alauzen et Lilya Slimani (2023). « Rules, Computation and Politics : Scrutinizing Unnoticed Programming Choices in French Housing Benefits ». In : *Journal of Cross-disciplinary Research in Computational Law* 1.3. (forthcoming). url : https://hal.inria.fr/hal-03712130.

Merigoux, Denis, Nicolas Chataing et Jonathan Protzenko (août 2021). « Catala : A Programming Language for the Law ». In : *Proc. ACM Program. Lang.* 5.ICFP. doi : 10.1145/3473582.

Merigoux, Denis, Raphaël Monat et Jonathan Protzenko (2021). « A Modern Compiler for the French Tax Code ». In : *Proceedings of the 30th ACM SIGPLAN International Conference on Compiler Construction*. CC 2021. Virtual, Republic of Korea : Association for Computing Machinery, p. 71-82. isbn : 9781450383257. doi : 10.1145/3446804.3446850.

Merigoux, Denis et Lilya Slimani (nov. 2022). *Literate programming snapshot of the Catala program for French housing benefits computation*. doi : 10.5281/zenodo.7357625. url : https://doi.org/10.5281/zenodo.7357625.

Ogien, Albert (1995). « L'esprit gestionnaire (une analyse de l'air du temps) ». In : *Recherches d'histoire et de sciences sociales*.

Olsen, Henrik Palmer, Jacob Livingston Slosser et Thomas Troels Hildebrandt (2020). « What's in the Box ? The Legal Requirement to Explain Computationally Aided Decision-Making in Public Administration ». In : *Constitutional Challenges in the Algorithmic Society. OUP*.

Pélisse, Jérôme (2019). « Varieties of legal intermediaries : when non-legal professionals act as legal intermediaries ». In : *Legal Intermediation : A Processual Approach to Law and Economic Activity*. Emerald Publishing Limited, p. 101-128.

Rosental, Claude (2019). *La société de démonstration*. Éditions du Croquant.

Roy, Marylou Le (mai 2023). « Un droit à l'explicabilité et à la "maîtrise" des algorithmes ». In : *Algorythms seminar*. url : https://www.seiller.org/Algorythms.html.

Siblot, Yasmine (2006). « Faire valoir ses droits au quotidien. Les services publics dans les quartiers populaires ». In : *Lectures, Les livres*.

Vasutiu, Ovidiu (2009). « Gestion des connaissances pour la maîtrise de la relation entre patrimoine documentaire et système d'information ». Thèse de doctorat dirigée par Pinon, Jean-Marie Informatique Lyon, INSA 2009. Thèse de doct. INSA Lyon, 1 vol. (X-204 p.) url : http://www.theses.fr/2009ISAL0100.

Vasutiu, Ovidiu, David Jouve et Youssef Amghar (jan. 2006). « Gestion des changements et étude d'impact dans un système d'information réglementaire. » In : p. 1007-1022.

Vessey, Iris (1985). « Expertise in debugging computer programs : A process analysis ». In : *International Journal of Man-Machine Studies* 23.5, p. 459-494.

Wachter, Sandra, Brent Mittelstadt et Chris Russell (2017). « Counterfactual explanations without opening the black box : Automated decisions and the GDPR ». In : *Harv. JL & Tech.* 31, p. 841.

Warin, Philippe (2016). « L'analyse du non-recours : au-delà du modèle de la relation de service ». FR. In : *Vie sociale* 14.2, p. 49-64. doi : 10.3917/vsoc.162.0049. url : https://www.cairn.info/revue-vie-sociale-2016-2-page-49.htm.

Weller, Jean-Marc (1999). *L'Etat au guichet : sociologie cognitive du travail et modernisation administrative des services publics*. Sociologie économique. Desclée de Brouwer, p. 254. url : https://halshs.archives-ouvertes.fr/halshs-00438938.

– (2003). « Le travail administratif, le droit et le principe de proximité ». FR. In : *L'Année sociologique* 53.2, p. 431-458. doi : 10.3917/anso.032.0431. url : https://www.cairn.info/revue-l-annee-sociologique-2003-2-page-431.htm.

Littérature grise

Association for Computing Machinery (2020). *Artifact Review and Badging Policy*. url : https://www.acm.org/publications/policies/artifact-review-and-badging-current.

Banuls, Justine et al. (2023). *La transparence des algorithmes publics*. Rapp. tech. Observatoire Data Publica. url : https://nextcloud.datactivist.coop/s/tqos5ppqGeEArDX.

CADA (2019). *Avis n°20181891*. https://www.cada.fr/20181891.

Changer de cap (2022). *Quand les algorithmes de la CAF ouvrent la chasse aux pauvres*. Silogora. url : https://silogora.org/quand-les-algorithmes-de-la-caf-ouvrent-la-chasse-aux-pauvres/.

CNAF (2023). *Réponse au collectif Changer de Cap*. url : https://changerdecap.net/wp-content/uploads/2023/03/02-01-Cnaf-reponse-au-Collectif-Changer-de-Cap.pdf.

Conseil d'État (2012). « Mme Labachiche épouse Beldjerrou ». In : *Section* 27 juillet 2012.n°347114.

Cour des Comptes (2023). *Certification des comptes 2022 du régime général de sécurité sociale et du CPSTI*. https://www.ccomptes.fr/fr/publications/certification-des-comptes-2022-du-regime-general-de-securite-sociale-et-du-cpsti.

Datactivist (2022). *Patchwork d'ouverture des algorithmes publics*. url : https://opendatacanvas.org/transparence-algo.

Donzel, Anne-Laure et al. (2022). « Expérimentation d'ouverture des algorithmes publics à la Métropole Européenne de Lille ». In : url : https://medium.com/datactivist/exp%C3%A9rimentation-douverture-des-algorithmes-publics-%C3%A0-la-m%C3%A9tropole-europ%C3%A9enne-de-lille-35c224053868.

Gaudin, Yann (2019). *Intermittents en fin de droits : réclamez votre dû!* https://blogs.mediapart.fr/yann-gaudin/blog/201119/intermittents-en-fin-de-droits-reclamez-votre-du.

Gnanasambandam, Chandra et al. (2023). « Yes, you can measure software developer productivity ». In : *McKinsey*. url : https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/yes-you-can-measure-software-developer-productivity.

Gosselin-Fleury, Geneviève et Damien Meslot (2013). *Mission d'information sur la mise en œuvre et le suivi de la réorganisation du ministère de la Défense*. url : https://www.assemblee-nationale.fr/14/rap-info/i1353.asp.

IEEE (2021). « IEEE Standard Model Process for Addressing Ethical Concerns during System Design ». In : *IEEE Std 7000-2021*, p. 1-82. doi : 10.1109/IEEESTD.2021.9536679.

La Quadrature du Net (2023a). *Notation des allocataires : l'indécence des pratiques de la CAF désormais indéniable*. https://www.laquadrature.net/2023/11/27/notation-des-allocataires-lindecence-des-pratiques-de-la-caf-desormais-indeniable/.

– (2023b). *Notation des allocataires : fébrile, la CAF s'enferme dans l'opacité*. https://www.laquadrature.net/2022/12/23/notation-des-allocataires-febrile-la-caf-senferme-dans-lopacite/.

Lequesne-Roth, Caroline, Mehdi Kimri et Pierre Legros (2021). *Livre blanc sur la digitalisation du service public : pour une éthique numérique inclusive*. url : https://www.fondationdenice.org/pour_une_ethique_numerique_inclusive/.

Ministère chargé de la ville et du logement (2023). *Les aides personnelles au logement : éléments de calcul*. https://www.ecologie.gouv.fr/sites/default/files/aides_personnelles_au_logement_elements_de_calcul_edition_2022.pdf.
Pénicaud, Soizic et Simon Chignard (2019). *Expliquer les algorithmes publics*. Etalab. url : https://guides.etalab.gouv.fr/algorithmes/guide.
Stallman, Richard, Roland Pesch, Stan Shebs et al. (1988). « Debugging with GDB ». In : *Free Software Foundation* 675.

Documents de presse

Berne, Xavier (2016). « Le fisc ouvrira le code source de son calculateur d'impôts le 1er avril ». In : *nextImpact*. url : https://www.nextinpact.com/article/21506/98981-le-fisc-ouvrira-code-source-son-calculateur-d-impots-1er-avril.
– (2018). « Les Allocations familiales nous ouvrent le code source de leur calculateur d'aides ». In : *Nextimpact*. url : https://www.nextinpact.com/article/28136/106298-les-allocations-familales-nous-ouvrent-code-source-leur-calculateur-daides.
Gauvin, Alice, Sarah Lerch et Marielle Krouk (2022). « La CAF ne répond pas ». In : *Envoyé Spécial*. url : https://www.francetvinfo.fr/replay-magazine/france-2/envoye-special/envoye-special-du-jeudi-1-decembre-2022_5470725.html.
Geiger, Gabriel (2021). « How a Discriminatory Algorithm Wrongly Accused Thousands of Families of Fraud ». In : *VICE*.
Gérard, Aline (2021). « Un Français sur trois ne détecterait pas une erreur sur sa fiche de paie ». In : *Ouest-France*. url : https://www.ouest-france.fr/economie/entreprises/un-francais-sur-trois-ne-detecterait-pas-une-erreur-sur-sa-fiche-de-paie-fbe7255a-921d-11eb-8153-111acea7321d.
Guillaud, Hubert (2018). « Vers des algorithmes exemplaires ? » In : *Internet Actu*.
Knaebel, Rachel (2022). « « Une galère pas possible » : quand la Caf refuse de prendre en compte la résidence alternée ». In : *Basta !* url : https://basta.media/RSA-APL-temoignage-une-galere-pas-possible-quand-la-caf-refuse-de-prendre-en-compte-la-residence-alternee.
Minot, Didier et Valérie Minot (2022). « Kafka à la CAF ». In : *Le 1*. url : le1hebdo.fr/journal/n-a/424/article/kafka--la-caf-5660.html.
Monin, Jacques (2018). *Louvois, le logiciel qui a mis l'armée à terre*. url : https://www.franceinter.fr/emissions/secrets-d-info/secrets-d-info-27-janvier-2018.
Romain, Manon et al. (2023). « Comment les algorithmes de la CAF prédisent si vous êtes à risque de frauder ». In : *Le Monde* 4 décembre 2023.
Zerouala, Faïza (19 juin 2021). « La réforme des APL vire au cauchemar pour les allocataires et ses agents ». In : *Mediapart*. url : https://www.mediapart.fr/journal/france/190621/la-reforme-des-apl-vire-au-cauchemar-pour-les-allocataires-et-ses-agents.

A Cas pratique juridique : calcul de l'aide au logement

Ce cas pratique vise à expliquer l'algorithme calculant, à la CNAF, le montant d'aide au logement versé dans le secteur locatif, à partir d'une situation fictive.

Madame X a introduit une demande d'aide au logement auprès de la CAF de Saint-Pierre (la collectivité d'outre-mer (COM) de Saint-Pierre-et-Miquelon) en date du 1er avril 2023.

Elle déclare être célibataire avec deux enfants à charge. Elle vit dans un appartement en location (conventionné) situé en zone II à Saint-Pierre-et-Miquelon. Elle dit payer 500 euros par mois et dispose de 12500 euros de ressources par an. Mme X ne bénéficie pas de la réduction de loyer de solidarité.

L'éligibilité de Mme X a été déterminée, celle-ci peut prétendre à l'allocation au logement.

Question : Considérant que Mme X est éligible à l'allocation au logement, de quel montant peut-elle bénéficier ?

Nota bene : Ce cas pratique ne remonte pas à la partie législative du Code de la construction et de l'habitation (CCH). En revanche, c'est un point important pour l'explicabilité technique du calcul puisqu'un élément de la calculette peut être défini par plusieurs articles. Quel(s) article(s) l'explication technique donne-t-elle ?

A.1 Définition du parc immobilier

L'article D. 823-9 du CCH liste les différentes situations ouvrant droit aux aides au logement en disposant que :

> Les modalités de liquidation et de versement des aides personnelles au logement sont fixées :
>
> 1. Pour les ménages occupant un logement dont ils sont locataires ou sous-locataires [...].

Le CCH définit ensuite les modalités de liquidation et de versement des aides personnelles au logement en fonction du parc immobilier (locatif, locatif foyer, accession à la propriété). Le barème du secteur locatif est défini par les articles D. 823-16 à 823-19 du CCH.

Toutefois, les aides au logement versées à Saint-Pierre-et-Miquelon font l'objet d'une réglementation spécifique prévue aux articles R. 863-1 du CCH. L'article D. 863-7 du CCH vient donc modifier le droit commun édifiant des barèmes différenciés selon le parc immobilier. En effet, cet article opère seulement une distinction entre l'accession à la propriété et le reste du parc immobilier (locatif et foyer). Il est précisé toutefois que pour le reste du parc immobilier, les articles D. 823-16 à 823-19 du CCH relatifs au barème du secteur locatif viendront s'appliquer.

En l'espèce, Mme X vit dans un appartement en location, conventionné, à Saint-Pierre-et-Miquelon.

Il convient de lui appliquer le 2° de l'article D. 823-9 du CCH tel que modifié par l'article D. 863-7 du CCH en ce qui concerne Saint-Pierre-et-Miquelon.

Il convient de lui appliquer le barème du secteur locatif.

A.2 Définition de la formule du calcul de l'aide

S'il existe trois types d'aides personnelles au logement (aide personnalisée au logement, allocation logement familiale et allocation logement sociale), la formule de ces aides est la même en ce qui concerne le secteur immobilier locatif. En effet, la section relative aux allocations logement (articles D. 842-1 à 842-4 du CCH) renvoie aux articles définissant le calcul de l'aide personnalisée au logement.

La formule du calcul des aides personnelles au logement du secteur locatif est définie aux articles D. 823-16 et suivants du CCH.

L'article D. 823-16 du CCH définie la formule du calcul du montant mensuel de l'*aide mensuelle* (Af). La formule est la suivante :

$$\mathrm{Af} = \mathrm{L} + \mathrm{C} - \mathrm{Pp}$$

Cet article définit ensuite les différents éléments du calcul :

1. « Af » est l'aide mensuelle issue de la formule de calcul ;
2. « L » est le loyer éligible, correspondant au loyer principal pris en compte dans la limite d'un plafond fixé par arrêté en fonction de la zone géographique et, sauf dans le cas où le logement occupé est une chambre, de la composition familiale ;
3. « C » est le montant forfaitaire au titre des charges, fixé par arrêté en fonction de la composition familiale ;
4. « Pp » est la participation personnelle du ménage calculée selon les dispositions de l'article D. 823-17.

Les cinq derniers aliénas de l'article décrivent des opérations de finalisation du calcul.

L'article suivant, D. 823-17 du CCH, détaille *la participation personnelle du ménage* (Pp). La formule de cet élément la suivante :

$$\mathrm{Pp} = \mathrm{P0} + \mathrm{Tp} \times (\mathrm{R} - \mathrm{R0})$$

L'article D. 823-17 dispose ainsi :

1. « Pp » est la participation personnelle du ménage ;
2. « P0 » est la participation minimale calculée selon des modalités précisées par arrêté et qui ne peut être inférieure à un montant minimum défini par arrêté ;
3. « Tp » est le taux de prise en compte des ressources du ménage. Il est égal à la somme d'un premier taux en fonction de la composition familiale et d'un second taux en fonction du loyer éligible défini au 2° de l'article D. 823-16. Le second taux est obtenu par l'application de taux progressifs à des tranches successives du loyer éligible, exprimé en proportion d'un loyer de référence en fonction de la composition familiale. Les valeurs du premier taux, les modalités de calcul du second taux et les valeurs des loyers de référence sont fixées par arrêté ;

4. « R » représente les ressources du ménage, appréciées selon les modalités prévues à la section 2 du chapitre II du présent titre et arrondies à la centaine d'euros supérieure ;
5. « R0 » est un abattement forfaitaire appliqué aux ressources du ménage. Il est fixé par arrêté en fonction de la composition familiale et est revalorisé au 1er janvier de chaque année, en fonction de l'évolution de l'indice des prix à la consommation des ménages hors tabac. Cette évolution est appréciée entre le 1er octobre de l'avant-dernière année précédant la revalorisation et le 1er octobre de l'année précédant la revalorisation. Il est arrondi à l'euro inférieur.

A.3 Définition du calcul de l'aide

Il est possible de diviser le calcul de l'aide en deux étapes, celle de la définition du montant global (A.3.1), puis celle de la finalisation du calcul (A.3.2).

A.3.1 Définition du montant

Les éléments du calcul sont attraits aux dépenses du logement, certains aux ressources. Puis, d'autres paramètres rentrent en compte dans le calcul.

Éléments relatifs au logement Au sein de l'article D. 823-16 du CCH précité, les éléments de calcul L (équivalence de loyer) et C (montant des charges) sont relatifs au logement. Au sein de l'article D. 823-17 du CCH ci-dessus, l'élément de calcul relatif au logement est la participation personnelle minimale (P0).

Équivalent de loyer L. L'article 7 de l'arrêté du 27 septembre 2019 en vigueur à compter du 1er juillet 2022 fixe des montants de plafond de loyers qui diffèrent selon la composition du foyer et de sa zone. Le plafond d'une personne seule avec deux enfants à charge en zone II est de 370,32 + 53,90 euros.

En l'espèce, la composition du ménage est d'une personne seule avec deux enfants à charge au 1er avril 2023. Le montant du loyer est de 500 euros.

$$\mathrm{L} = 370{,}32 + 53{,}90 = 424{,}22 \Rightarrow \mathrm{L} < 500$$

En conclusion, l'équivalent loyer L est de 424,22 euros.

Montant des charges C. L'article 9 de l'arrêté du 27 septembre 2019 en vigueur à compter du 1er juillet 2022 fixe des montants forfaitaires des charges (C) à 56,12 euros pour un bénéficiaire isolé ou un couple sans personne à charge auquel s'ajoutent 12,72 euros par personne supplémentaire à charge.

En l'espèce, la composition du ménage est de une personne seule avec deux enfants à charge au 1er avril 2023.

$$\mathrm{C} = 56{,}12 + 12{,}72 \times 2 = 81{,}56$$

En conclusion, le montant forfaitaire des charges est égal à 81,56 euros.

Participation personnelle minimale P0. L'article 13 de l'arrêté du 27 septembre 2019 en vigueur à compter du 1er juillet 2022 dispose ainsi que la P0 : « est égale à la plus élevée des deux valeurs suivantes : 8,5% de la somme du loyer éligible défini au 2° de l'article D. 823-16 du même code et du forfait charge ou 36,63 euros." »

En l'espèce, C et L déterminés auparavant sont respectivement égaux à 81,56 et 424,22.

Il convient donc de faire comparer la somme de C et L multiplié par 8,5% à 36,63 euros, afin de déterminer le montant le plus élevé.

$$(\mathrm{C} + \mathrm{L}) \times 8{,}5\% = 505{,}78 \times 8{,}5\% = 42{,}99 > 36{,}63 \Rightarrow \mathrm{P0} = 42{,}99$$

En conclusion, la P0 est égale à 42,99 euros.

Éléments relatifs aux ressources Les modalités générales de l'appréciation des ressources sont précisées aux articles R822-2 du CCH et suivants.

Ressources du ménage R. L'article D. 823-17 énonce au 4° que les ressources prises en compte doivent être arrondies à la centaine d'euros supérieure.

En l'espèce, Mme X dispose de 12500 euros de ressources par an.

Les ressources de Mme X tombent déjà sur une centaine d'euros. Il n'y a donc pas besoin d'arrondir les ressources.

$$\mathrm{R} = 12500$$

En conclusion, les ressources R sont égales à 12500 euros.

Abattement forfaitaire R0. En application du 5° de l'article D. 823-17 du CCH, l'article 15 en vigueur de l'arrêté du 27 septembre 2019 décline les montants de l'abattement forfaitaire R0 pour les prestations dues à compter du 1er janvier 2023. Des montants différents sont applicables à Saint-Pierre-et-Miquelon, définis à l'article 47 de cet arrêté. Les montants sont fonction de la composition du foyer.

En l'espèce, le 1er avril, Mme X qui habite seule à Saint-Pierre-et-Miquelon a deux enfants à charge.

Le montant attribué à une personne seule avec deux enfants à charge pour les prestations dues à compter du 1er janvier 2023 est de 9211 euros.

$$\mathrm{R0} = 9211$$

En conclusion, l'abattement R0 est égal à 9211 euros.

Autres paramètres *Taux de participation personnel (Tp)*. En application du 3° de l'article D. 823-17 du CCH, l'article 14, en vigueur depuis le 1er juillet 2022, de l'arrêté du 27 septembre 2019 détaille la formule pour obtenir le Tp. La formule est la suivante :

$$\mathrm{Tp} = \mathrm{TF} + \mathrm{TL}$$

Tp est composé de la somme de deux taux, le taux fonction de la taille de la famille (TF) et le taux complémentaire (TL).

Taux fonction de la taille de la famille (TF). Selon le 1° de l'article 14, le TF est fixe, fonction de la taille de la famille.

En l'espèce, la composition familiale de Mme X est une personne seule avec deux personnes à charge.

Il convient de lui appliquer le taux défini pour une personne seule ayant deux personnes à charge, soit 2,38%.

$$\mathrm{TF} = 2{,}38\%$$

Taux complémentaire (TL). Le 2° de l'article 14 précité dispose ainsi :

> L représente un taux complémentaire fixé ci-dessous en fonction de la valeur du rapport RL entre le loyer retenu dans la limite du plafond L et un loyer de référence LR.
>
> $$\mathrm{RL} = \frac{\mathrm{L}}{\mathrm{LR}}$$
>
> RL est exprimé en pourcentage et arrondi à la deuxième décimale.

Loyer de référence (LR). Le loyer de référence LR est défini au 2°. Il est fixe et fonction de la composition du foyer. Pour une personne seule ayant une personne à charge, le montant est de 370,32 euros, et 53,90 euros par personne à charge supplémentaire.

En l'espèce, le L a été déterminé ci-dessus et est égal à 424,22 euros. Mme X est une personne seule ayant deux personnes à charge.

$$\mathrm{LR} = 370{,}32 + 53{,}90 \times 1 = 424{,}22$$

D'où :

$$\mathrm{RL} = \frac{424{,}22}{424{,}22} = 100\%$$

Le RL est égal à 100%.

Le 2° de l'article 14 définit des formules avec des taux progressives selon le RL. Pour les RL supérieurs à 75%, la formule est la suivante, arrondi à la troisième décimale :

$$\mathrm{TL} = 0{,}45\% \times 30\% + 0{,}68\% \times (\mathrm{RL} - 75\%)$$

En l'espèce, RL = 100%.

$$
\begin{aligned}
&\mathrm{RL} > 75\% \\
&\mathrm{TL} = 0{,}45\% \times 30\% + 0{,}68\% \times (100\% - 75\%) \\
&\mathrm{TL} = 0{,}00135 + 0{,}68\% \times 0{,}25 \\
&\mathrm{TL} = 0{,}00135 + 0{,}00171 \\
&\mathrm{TL} = 0{,}00305 \\
&\mathrm{TL} = 0{,}305\%
\end{aligned}
$$

Le TL est égal à 0,305%.

Il convient désormais d'additionner le TF et le TL pour obtenir le Tp.

$$\mathrm{Tp} = 2{,}38\% + 0{,}305\% = 2{,}685\%$$

En conclusion, le Tp est égal à 2,685%.

Dorénavant, tous les éléments ont été déterminés pour calculer le montant de la Pp.

Pour rappel, la formule du montant de la Pp est la suivante :

$$\mathrm{Pp} = \mathrm{P0} + \mathrm{Tp} \times (\mathrm{R} - \mathrm{R0})$$

En l'espèce, la P0 = 42,99, le Tp = 2,685%, les R = 12500 et le R0 = 9211.

$$
\begin{aligned}
&\mathrm{Pp} = 42{,}99 + 2{,}685\%(12500 - 9211) \\
&\mathrm{Pp} = 42{,}99 + 2{,}685\% \times 3289 \\
&\mathrm{Pp} = 42{,}99 + 88{,}31 \\
&\mathrm{Pp} = 131{,}30
\end{aligned}
$$

En conclusion, la Pp est égale à 131,30 euros.

Dorénavant, tous les éléments ont été déterminés pour caculer le montant de l'aide.

Pour rappel, la formule du montant de l'aide est la suivante :

$$\mathrm{Af} = \mathrm{L} + \mathrm{C} - \mathrm{Pp}$$

En l'espèce, le L = 424,22 et les C = 81,56 et la Pp = 131,30.

$$\mathrm{Af} = 424{,}22 + 81{,}56 - 131{,}30 = 374{,}48$$

En conclusion, l'Af est égale à 374,48 euros.

A.3.2 Finalisation du calcul

Les 5 derniers aliénas de l'article D. 823-16 du CCH prévoit des opérations de finalisation du calcul.

Dégressivité et suppression Selon l'aliéna 8 de l'article D. 823-16 du CCH, le montant calculé peut être diminué ou supprimé dans le cas où le loyer dépasse un plafond obtenu par l'application de coefficients multiplicateurs au plafond de loyer fixé au 2° de ce même article du CCH. L'article 10 de l'arrêté de 2019 fixe des coefficients multiplicateurs qui sont fonction de la zone géographique. Celui-ci est fixé à 2,5 (diminution) et à 3,1 (suppression) pour la zone II.

En l'espèce, le plafond de loyer L prédéterminé est égal à 424,22. Mme X habite dans la zone II et son loyer est de 500 euros.

$$424{,}22 \times 2{,}5 = 1060{,}55 > 500$$

En conclusion, il n'y a ni dégressivité ni suppression du montant de l'aide à raison de l'alinéa 8 de l'article D. 823-16 du CCH.

Minoration forfaitaire L'article 11 de l'arrêté de 2019 précité fixe à 5 euros la minoration forfaitaire.

En l'espèce, l'AF, avant les opérations de finalisation du calcul est égale à 374,48 euros.

$$374{,}48 - 5 = 369{,}48$$

En conclusion, le montant après la minoration forfaitaire est égal à 369,48 euros.

Contribution au remboursement de la dette sociale Comme mentionné à l'article D. 823-16 du CCH, l'ordonnance du 24 janvier 1996 relative au remboursement de la dette sociale prévoit la CRDS à son article 14 en vigueur le 1er septembre 2018 sur les aides personnelles au logement. L'article 19 de cette même ordonnance fixe un taux à 0,5%. Selon l'alinéa 10 de l'article D. 823-16 du CCH :

> Le montant [...] est diminué d'un montant représentatif des contributions sociales qui s'y appliquent, arrondi à l'euro inférieur, puis majoré de ce montant représentatif.

Toutefois, l'article LO 6413-1 du CGCT rappelle le régime constitutionnel des collectivités territoriales d'outre-mer régies par l'article 74 de la Constitution, qui dispose que :

> Les dispositions législatives et réglementaires sont applicables de plein droit à Saint-Pierre-et-Miquelon, à l'exception de celles qui interviennent dans les matières relevant de la loi organique en application de l'article 74 de la Constitution ou dans l'une des matières relevant de la compétence de la collectivité en application du II de l'article LO 6414-1.

De plus, l'archipel de SPM dispose de la compétence en matière d'impôts en vertu de l'article LO6414-1 du Code général des collectivités territoriales (CGCT).

Ainsi, en l'absence de dispositions particulières instituant la CRDS à Saint-Pierre-et-Miquelon, la CRDS n'est pas un impôt dédié prélevé sur la collectivité d'outre-mer. En revanche, selon l'alinéa 10 de l'article D. 823-16 du CCH, le montant est arrondi à l'euro inférieur.

En l'espèce, le montant après la minoration forfaitaire est égal à 369,48 euros. Arrondi à l'euro inférieur, il est égal à 369 euros.

En conclusion, le montant est égal à 369 euros.

Réduction de loyer de solidarité L'alinéa 11 de l'article D. 823-16 du CCH prévoit une déduction du montant pour les personnes touchant la réduction de loyer de solidarité.

En l'espèce, Mme X ne bénéficie pas de la réduction de loyer de solidarité.

En conclusion, le montant n'est pas diminué.

Montant minimal L'article 12 de l'arrêté de 2019 précité fixe un seuil de versement avant application de la contribution au remboursement de la dette sociale (CRDS). Celui-ci est fixé à 10 euros pour les allocations de logement.

En l'espèce, le montant avant application de la CRDS est de 369,48 euros.

$$369{,}48 > 10$$

En conclusion, le montant peut être versé.

Montée en charge des aides pour SPM Les aides au logement ont été établies à SPM par le décret du 21 décembre 2021 portant diverses mesures sur les aides personnelles au logement et relatif aux aides personnelles au logement à Saint-Pierre-et-Miquelon. Le décret instaure une montée en charge des aides progressive jusqu'à 2026. Pour obtenir l'aide finale à verser, il convient d'appliquer la formule suivante à compte du mois de janvier 2022 jusqu'au mois de décembre 2025 :

$$\text{ALspm} = \text{ALt} \times \left(1 - \frac{2026 - \text{n}}{8}\right)$$

Dans laquelle :

1. « ALspm » est l'aide mensuelle due aux bénéficiaires de l'aide personnelle au logement à Saint-Pierre-et-Miquelon au titre du mois considéré, entre janvier 2022 et décembre 2025 ;
2. « ALt » est l'aide mensuelle théorique calculée en application des dispositions du livre VIII du code précité, le cas échéant adaptées selon les dispositions de l'article 2 du présent décret ;
3. « n » correspond à l'année au titre de laquelle l'aide est due. Les dispositions du présent article peuvent être modifiées par décret.

En l'espèce, Mme X vit à Saint-Pierre-et-Miquelon le 1er avril 2022 date à laquelle elle a effectué sa demande. De plus, l'aide finale est égale à 369,48 euros.

$$\begin{aligned}\text{ALspm} &= 369 \times \left(1 - \frac{2026 - 2023}{8}\right)\\ \text{ALspm} &= 369 \times (1 - 0{,}375)\\ \text{ALspm} &= 369 \times 0{,}625\\ \text{ALspm} &= 230{,}63\end{aligned}$$

À titre conclusif, le montant à verser à Mme X est de 230,63 euros.

B Sémantique d'évaluation paresseuse pour le calcul par défaut

Cette annexe est destinée aux lectrices familières de la science des langages de programmation et de leur formalisme, car elle est présentée dans le style habituel des articles de recherche de ce domaine de recherche. Elle reprend les notations de l'article [Merigoux, Chataing et Protzenko 2021] dont elle est en quelque sorte une extension.

B.1 Sémantique du calcul par défaut étendue

Type	τ_0	::=	`bool` | `unit`	types booléen et unité
		|	$\tau \to \tau$	type fonction
	τ	::=	τ_0	
		|	$\{\tau_0\}$	type résultat de défaut
Expression	e	::=	x | s | `true` | `false` | `()`	variable, nom de décl., littéraux
		|	`op`$(e, \ldots, e)$	application d'opérateurs prédéfinis
		|	$\lambda\ (x : \tau)\ .\ e$ | $e\ e$	λ-calcul
		|	d | `Err`	
Défauts	d	::=	$\langle e \mid e :\!- e \rangle$	terme défaut
		|	`check-empty`(e)	extraction de résultat
		|	$\{e\}$	défaut non vide
		|	$\varnothing$	défaut vide
Erreurs	`Err`	::=	$\texttt{Err}^{\varnothing}$	terme d'erreur (vide)
		|	$\texttt{Err}^{\circledast}$	terme d'erreur (conflit)

Figure 13 – Calcul par défaut avec valeurs de retours explicites et termes d'erreur

Nous partons de la sémantique de Catala définie dans Merigoux, Chataing et Protzenko [2021], que nous avons étendue pour différencier les termes vides dans le contexte où ils sont valides (à l'intérieur des termes par défaut), des termes vides d'erreur correspondant à un échec du programme (voir Figure 13). Dans ce but, une nouvelle construction $\{e\}$ correspondant au résultat non-vide de l'évaluation du terme $\langle \varnothing \mid \texttt{true} :\!- e \rangle$ est ajoutée. L'opération inverse `check-empty`(e) permet de lever explicitement l'erreur là où un résultat non vide est attendu. Ces ajouts nous permettent de raffiner les règles de typage comme le montre la Figure 14 :

nous introduisons un nouveau type $\{\tau_0\}$ qui signale une valeur de type τ_0 qui peut parfois être s'évaluer vers le terme vide d'erreur.

$$\text{T-Error}\quad \Gamma \vdash \texttt{Err} : \tau \qquad \text{T-Empty}\quad \Gamma \vdash \varnothing : \{\tau\}$$

$$\text{T-DefaultRet}\quad \frac{\Gamma \vdash e : \tau}{\Gamma \vdash \{e\} : \{\tau\}} \qquad \text{T-CheckEmpty}\quad \frac{\Gamma \vdash e : \{\tau\}}{\Gamma \vdash \texttt{check-empty}(e) : \tau}$$

$$\text{T-Default}\quad \frac{\Gamma \vdash \{e_i\} : \tau \qquad \Gamma \vdash e_{\text{just}} : \texttt{bool} \qquad \Gamma \vdash e_{\text{cons}} : \{\tau\}}{\Gamma \vdash \langle e_1, \ldots, e_n \mid e_{\text{just}} :\!- e_{\text{cons}} \rangle : \{\tau\}}$$

Figure 14 – Typing rules for the default calculus

B.2 Sémantique paresseuse et évaluation partielle

Afin de définir la sémantique d'évaluation partielle que nous recherchons ici, nous allons raffiner la notion de valeurs, dont il existera plusieurs variantes suivant le contexte, comme montré en Figure 15. Tout d'abord, les valeurs atomiques qui reprennent les valeurs précédentes de la sémantique de Merigoux, Chataing et Protzenko [2021]. Ensuite, les valeurs d'opérations qui stockent en réalité des valeurs partiellement évaluées, puisqu'elles contiennent l'arbre de syntaxe des opérateurs à appliquer pour calculer le résultat final. Ces valeurs d'opérations vont nous permettre d'expliquer le résultat final du calcul en montrant quelles opérations ont été appliquées pour y arriver. Enfin, les valeurs de défaut correspondent aux termes temporaires créés par notre sémantique paresseuse lors de l'évaluation des défauts. En effet, la valeur $\{e\}$ signifie que pendant l'évaluation des exceptions d'un défaut, nous pouvons considérer qu'un terme est non-vide sans en évaluer complètement l'expression e.

Valeurs atomiques	v	::=	$\lambda\,(x : \tau)\,.\,e$	fonctions
		|	$\texttt{true} \mid \texttt{false}$	booléens
		|	$\texttt{Err}^{\varnothing} \mid \texttt{Err}^{\circledast}$	erreurs
		|	$\{v\} \mid \varnothing$	défauts
Valeurs d'opérations	v_{op}	::=	$v \mid \texttt{op}(v_{\text{op}}, \ldots, v_{\text{op}})$	
Valeurs de défaut	v_{dft}	::=	$\varnothing$	terme vide
		|	$\{e\}$	terme non-vide

Figure 15 – Différents niveaux de valeurs

La Figure 16 présente les règles de notre nouvelle évaluation paresseuse et partielle définie par la réduction $\longrightarrow$. Elle est très proche de la réduction originelle définie par Merigoux, Chataing et Protzenko [2021], à deux exceptions près.

Premièrement, il n'y a pas de règles pour évaluer l'application des opérateurs. Ceci permet d'obtenir l'évaluation partielle puisqu'on ne peut pas réduire $\texttt{op}(v_{\text{op}}, \ldots, v_{\text{op}})$ qui est une valeur d'opérations. Deuxièmement et cependant, il est nécessaire d'effectuer une réduction plus complète pour les termes qui influent sur le flot de contrôle du programme. C'est le cas à deux endroits : les exceptions des défauts et les justifications des défauts. Pour les exceptions des défauts, il suffit de savoir si le terme réduit strictement vers $\varnothing$ ou pas. Par un argument de typage, on peut avancer que $\varnothing$ apparaît soit tout seul soit après la réduction

D-Beta
$(\lambda\ (x : \tau)\ .\ e)\ v_{\text{op}} \longrightarrow e[x \mapsto v_{\text{op}}]$

D-ErrorPropagation
$$\frac{e \longrightarrow \texttt{Err}}{C[e] \longrightarrow \texttt{Err}}$$

D-Context
$$\frac{e \longrightarrow e' \qquad e' \neq \texttt{Err}}{C[e] \longrightarrow C[e']}$$

D-CheckEmptyOk
$\texttt{check-empty}(\{e\}) \longrightarrow e$

D-CheckEmptyErr
$\texttt{check-empty}(\varnothing) \longrightarrow \texttt{Err}^{\varnothing}$

D-DefaultTrueNoException
$\langle \varnothing, \ldots, \varnothing \mid \texttt{true} :- e \rangle \longrightarrow e$

D-DefaultFalseNoException
$\langle \varnothing, \ldots, \varnothing \mid \texttt{false} :- e \rangle \longrightarrow \varnothing$

D-DefaultOneException
$\langle \varnothing, \ldots, \varnothing, \{e\}, \varnothing, \ldots, \varnothing \mid e_1 :- e_2 \rangle \longrightarrow \{e\}$

D-DefaultTooManyExceptions
$\langle \ldots, \{e_1\}, \ldots, \{e_2\}, \ldots \mid e_3 :- e_4 \rangle \longrightarrow \texttt{Err}^{\circledast}$

Figure 16 – Règles de réduction

$$\begin{array}{rcll} C & ::= & \cdot\ e \mid v_{\text{op}}\ \cdot & \text{évaluation des applications} \\ & \mid & \langle \vec{v_{\text{dft}}}, \cdot, \vec{e} \mid e :- e \rangle & \text{évaluation des exceptions} \end{array}$$

Figure 17 – Contextes d'évaluation

d'un défaut, et donc les règles de la Figure 16 suffisent pour assurer une sémantique stricte de réducation pour cet objectif. Le contexte $\{\cdot\}$ a volontairement été omis de la Figure 17, ce qui matérialise le caractère paresseux de l'évaluation des exceptions des défauts.

Par contre, pour les justifications des défauts, il est nécessaire d'évaluer de manière plus complète pour obtenir une valeur booléenne qui détermine le résultat du défaut. Un nouveau contexte correspondant à l'évaluation déterminant le flot de contrôle du programme, C_{control}, défini en Figure 18 permet d'achever ces résultats. Si le langage possédait une construction conditionnelle classique, l'évaluation de la condition booléenne apparaîtrait aussi dans C_{control}. Dans ce contexte, les règles de réduction précédentes s'appliquent toujours, mais on ajoute la règle D-EvaluateOperator qui effectue l'évaluation concrète des opérateurs et produit le résultat concret du calcul sous la forme de valeur atomique.

$$\begin{array}{rcll} C_{\text{control}} & ::= & \langle \vec{v_{\text{dft}}} \mid \cdot :- e \rangle & \text{évaluation des justifications des défauts} \\ & \mid & \texttt{op}(\vec{v_{\text{op}}}, \cdot, \vec{e}) & \text{évaluation des arguments des opérateurs} \end{array}$$

D-ContextControl
$$\frac{e \longrightarrow e' \qquad e' \neq \texttt{Err}}{C_{\text{control}}[e] \longrightarrow C_{\text{control}}[e']}$$

D-ErrorPropagationControl
$$\frac{e \longrightarrow \texttt{Err}}{C_{\text{control}}[e] \longrightarrow \texttt{Err}}$$

D-OperatorEvaluationControl
$$\frac{e \longrightarrow e'}{C_{\text{control}}[e] \longrightarrow C_{\text{control}}[e']}$$

D-EvaluateOperator
$$\frac{\text{operation result is } v}{\texttt{op}(v_1, \ldots, v_n) \longrightarrow v}$$

Figure 18 – Réduction complète des expressions déterminant le flot de contrôle du programme

B.3 Cohérence de la sémantique d'évaluation partielle

Nous énonçons maintenant quelques théorèmes de cohérence que cette formalisation devrait satisfaire, dont nous laissons les preuves à des travaux futurs. Dans ces énoncés, on fait référence à la réduction stricte de Merigoux, Chataing et Protzenko [2021] par $\xrightarrow{\text{eager}}$, en supposant qu'elle a été adaptée aux extensions de typage de B.1.

Théorème 1 (Sûreté du typage) *Pour toute expression bien typée $e : \tau$, alors soit e est une valeur v_{op} ; soit $e \longrightarrow e' : \tau$.*

Théorème 2 (Équivalence des sémantiques) *Soit une expression bien typée e, τ. On a les deux sens suivants de l'équivalence :*

- *s'il existe v_{op} telle que $e \longrightarrow^* v_{op}$, alors il existe v telle que $e \xrightarrow{\text{eager}}{}^* v$ et soit $v = \texttt{Err}$, soit $v_{op} \xrightarrow{\text{eager}}{}^* v$;*
- *s'il existe v telle que $e \xrightarrow{\text{eager}}{}^* v$, alors il existe v_{op} telle que $e \longrightarrow^* v_{op}$ et $v_{op} \xrightarrow{\text{eager}}{}^* v$.*

CENTRE DE RECHERCHE DE PARIS

2 rue Simone Iff - CS 42112
75589 Paris Cedex 12

Éditeur
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399