



D

RAPPORT

Algorithmes, systèmes d'IA et services publics : quels droits pour les usagers ?

POINTS DE VIGILANCE ET RECOMMANDATIONS

Pour que le droit n'oublie personne

Défenseur des droits

RÉPUBLIQUE FRANÇAISE

RAPPORT

Algorithmes, systèmes d'IA et services publics :
quels droits pour les usagers ?

Points de vigilance et recommandations

PROPOS LIMINAIRES

Alors que les algorithmes et les systèmes d'intelligence artificielle (IA) transforment à la fois les services publics et le travail des agents publics, la Défenseure des droits a jugé nécessaire, en tant qu'institution en charge de la défense des droits des usagers des services publics, d'investir ce sujet.

En miroir de la place prise par les algorithmes et les systèmes d'IA dans nos vies, une littérature foisonnante est aujourd'hui consacrée à leurs développements techniques et juridiques ainsi qu'aux enjeux que ceux-ci soulèvent en termes éthiques, juridiques, environnementaux, sociétaux et commerciaux.

Le Défenseur des droits a, dans la lignée des rapports dédiés à la dématérialisation des services publics¹, commencé à contribuer au nécessaire débat public sur ces outils. S'ils sont porteurs de progrès, les systèmes algorithmiques font aussi peser des risques majeurs sur les droits et libertés, comme l'Institution l'a souligné dans deux rapports consacrés, l'un en 2020, au risque d'automatisation des discriminations induit par les algorithmes² et, l'autre en 2021, à l'impératif respect des droits fondamentaux dans l'utilisation des technologies biométriques³.

Alors que les algorithmes et les systèmes d'IA sont déployés de plus en plus massivement dans les politiques publiques et les services publics, l'action publique comme les droits des usagers concernés s'en trouvent substantiellement affectés. De plus en plus de décisions administratives individuelles sont partiellement ou entièrement automatisées, prises sur la base de résultats livrés par des algorithmes ou des systèmes d'IA.

Les réclamations reçues par l'institution sur ces sujets restent à ce jour peu nombreuses : les réclamants contestent davantage le résultat des décisions dont ils sont l'objet que le processus qui y a conduit. En effet, les modalités de prise de décision, dont le caractère automatisé et le fonctionnement,

voire les éventuels biais, ne sont pas visibles de façon directe. Mais certaines réclamations reçues donnent à voir des difficultés qui apparaissent potentiellement systémiques, notamment sur la procédure Affelnet.

À cet égard, des décisions de justice ainsi que des travaux de doctrine juridique et de recherche scientifique mettent en évidence des problématiques persistantes, et ce dans le sillage de nombreuses initiatives conduites par la Commission nationale Informatique et Libertés (Cnil)⁴, le Conseil d'État⁵ et d'autres institutions publiques⁶ ou privées de recherche ou de formation⁷. Sont en particulier interrogées la substance de l'intervention humaine dans les cas où la décision administrative individuelle n'est dite que « partiellement automatisée » ainsi que la nature des explications données à l'utilisateur concerné sur le fonctionnement de l'algorithme ou du système d'IA grâce auquel la décision a été - entièrement ou non - prise.

C'est sur la base de ces savoirs et en s'appuyant sur des constats et propositions déjà formulées que la Défenseure des droits a souhaité consacrer un rapport à ces enjeux. Il s'agit aussi de réaffirmer la vision d'un service public devant toujours conserver son rôle de soutien et de service au public, par le respect des grands principes de continuité du service, d'égalité de tous devant le service et d'adaptabilité aux besoins des usagers. L'institution promeut ainsi dans l'ensemble des administrations une véritable « culture des droits », centrée sur les usagers. Cela suppose de placer le respect des droits des usagers au cœur des objectifs assignés aux managers publics comme aux agents de terrain et, corrélativement, de ne pas laisser l'administration se concentrer, notamment s'agissant du déploiement des systèmes algorithmiques, sur des objectifs prioritaires de réduction des coûts et de personnel et sur des indicateurs de rentabilité.

La place croissante des outils algorithmiques dans l'action publique se nourrit de la dématérialisation des démarches administratives. S'agissant des droits des usagers, examiner les enjeux de l'action publique algorithmique ne doit pas conduire à occulter l'importance des problématiques spécifiques liées à la « seule » dématérialisation des services publics. Le rapport annuel d'activité 2023 du Défenseur des droits souligne que l'accès aux droits pouvait être mis à mal notamment par la dématérialisation qui met à l'écart dix millions de personnes⁸. Il rappelle aussi que la dématérialisation des procédures administratives doit être une offre complémentaire, non substitutive, au guichet, au courrier papier et au téléphone, afin de garantir un accès équitable aux services publics⁹. L'institution continuera donc de veiller au suivi des recommandations qu'elle a émises à cet égard¹⁰.

Si le cadre juridique encadrant l'usage des algorithmes et systèmes d'IA évolue, les règles suivantes s'appliquent, de façon cumulative ou non :

- Dès lors qu'il y a traitement de données personnelles, les obligations et garanties prévues par la loi « Informatique et Libertés »¹¹ (régulièrement modifiée, notamment en 2018 pour l'adapter au règlement général sur la protection des données, dit RGPD¹²), pour lesquelles la Cnil est compétente ;
- Les obligations prévues en 2016 par la loi pour une République numérique¹³, désormais codifiées au sein du code des relations entre le public et l'administration (CRPA), dès qu'il y a utilisation par une administration d'un algorithme pour prendre ou aider à prendre une décision administrative individuelle, qu'il y ait ou non traitement de données personnelles ;
- Les obligations spécifiques, liées à certains secteurs comme par exemple dans le domaine de l'accès à l'enseignement supérieur¹⁴.

En outre, les obligations issues du règlement sur l'IA adopté par l'Union européenne en 2024 (*AI Act* pour *Artificial intelligence Act*)¹⁵ pour encadrer le développement, la mise sur le marché et l'utilisation de systèmes d'IA, commenceront à s'appliquer dès 2025 et ceci de façon progressive. En 2030, seront applicables toutes les règles relatives aux systèmes d'IA à haut risque listés à l'annexe III du texte¹⁶ et notamment les systèmes d'IA utilisés dans les domaines de l'éducation, de l'emploi, de l'accès aux services publics essentiels, de l'application de la loi, de l'immigration et de l'administration de la justice.

Enfin, le Conseil de l'Europe a adopté à l'été 2024 une convention-cadre sur l'intelligence artificielle et les droits de l'Homme, la démocratie et l'État de droit¹⁷, que la Commission européenne a signé en septembre 2024 au nom de l'Union européenne.

Le présent rapport, qui s'adresse à la fois au gouvernement, au Parlement, aux responsables publics (administrations, collectivités), aux associations d'accès aux droits et aux usagers eux-mêmes, ne se donne pas pour objectif de traiter de tous les enjeux posés par le déploiement, dans les services publics, des algorithmes et des systèmes d'IA, et de l'ensemble des risques associés – notamment discriminatoires –, qui feront l'objet de travaux ultérieurs. Il s'intéresse spécifiquement aux décisions prises avec l'appui de systèmes algorithmiques et d'IA et aux problématiques conséquentes, d'une part en matière d'intervention humaine et d'autre part s'agissant de la transparence et de l'explicabilité de ces décisions. Il entend être une contribution aux débats qui doivent continuer à se tenir en vue du respect des droits d'ores et déjà applicables, ou de ceux qui devraient l'être, pour les usagers des services publics.

TABLE DES MATIÈRES

| | | | |
|--|-----------|--|-----------|
| I· LE DÉPLOIEMENT DES ALGORITHMES ET DES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE (IA) PAR L'ADMINISTRATION DANS SES RELATIONS AVEC LE PUBLIC | 6 | III· LA TRANSPARENCE : UN PRINCIPE FORT, MAIS EN PRATIQUE ? | 28 |
| 1· DÉFINITIONS PRÉLIMINAIRES | 6 | 1· DES ENJEUX MAJEURS | 29 |
| 2· L'ALGORITHMISATION DE L'ADMINISTRATION | 8 | 2· DES DROITS D'ORES ET DÉJÀ EN VIGUEUR | 32 |
| 3· L'OUVERTURE À LA PRISE DE DÉCISIONS ADMINISTRATIVES INDIVIDUELLES ENTIÈREMENT AUTOMATISÉES | 10 | 3· DES LIMITES JURIDIQUES, UNE RECOMMANDATION POUR DES AMÉLIORATIONS | 36 |
| | | 4· DES LIMITES PRATIQUES ET DES RECOMMANDATIONS POUR Y REMÉDIER | 37 |
| II· LA MAÎTRISE ET L'INTERVENTION HUMAINE : DES GARANTIES, MAIS CONCRÈTEMENT ? | 14 | | |
| 1· LES DIFFÉRENTS NIVEAUX DE L'INTERVENTION HUMAINE | 14 | | |
| Les contrôles « système » | 15 | | |
| Les contrôles « individuels » | 21 | | |
| 2· LA SUBSTANCE DE L'INTERVENTION HUMAINE AU MOMENT DE LA PRISE DE DÉCISION INDIVIDUELLE – RECOMMANDATIONS POUR LA GARANTIR | 21 | | |

PARTIE I

LE DÉPLOIEMENT DES ALGORITHMES ET DES SYSTÈMES D'INTELLIGENCE ARTIFICIELLE (IA) PAR L'ADMINISTRATION DANS SES RELATIONS AVEC LE PUBLIC

Quelques éléments de définition s'imposent avant de dresser le rapide panorama de l'utilisation de plus en plus généralisée d'algorithmes par l'administration (« l'algorithmisation de l'administration ») et, en parallèle, de l'ouverture progressive à la prise de décisions administratives individuelles entièrement automatisées.

1 · DÉFINITIONS PRÉLIMINAIRES

L'algorithme ne fait l'objet **d'aucune définition légale** mais en s'appuyant sur les travaux de la Cnil et de la Commission d'accès aux documents administratifs (Cada), on peut définir **un algorithme** comme « *la description d'une suite finie et non ambiguë d'étapes (ou d'instructions) permettant d'obtenir un résultat à partir d'éléments fournis en entrée. Par exemple, une feuille de calcul permettant de déterminer un droit à une prestation à partir de critères prédéfinis constitue un algorithme* »¹⁸. Il faut préciser que : « *Pour qu'un algorithme puisse être mis en œuvre par un ordinateur, il faut qu'il soit exprimé dans un langage informatique, transcrit en un programme (texte composé de commandes écrites, également appelé "code source")* ».

Ce programme peut être alors exécuté dans un logiciel ou compilé sous la forme d'une application. Un logiciel a en général recours à de nombreux algorithmes : pour la saisie des données, le calcul du résultat, leur affichage, la communication avec d'autres logiciels, etc. »¹⁹. C'est, en fait, une sorte de recette de cuisine : le système prend des données en entrée (les ingrédients), sur lesquelles il effectue des opérations (dosage, mélange, cuisson, etc.), afin d'obtenir un résultat en sortie (un plat).

Des algorithmes sont utilisés depuis des décennies. Par exemple, l'algorithme utilisé par la Direction générale des finances publiques (DGFIP) pour calculer l'impôt sur le revenu²⁰.

La loi et le règlement posent des obligations pour les organismes et notamment pour les administrations qui utilisent des algorithmes pour fonder des décisions individuelles.

De nouveaux enjeux se déploient avec la troisième révolution industrielle et le *big data*. Le *big data* recouvre l'ensemble des données que les personnes génèrent dans leur quotidien par l'utilisation d'outils numériques (messages, signaux GPS, photos, données transactionnelles...). Ces données, créées par et à propos des individus, regroupent des données volontaires, des données observées (nos traces numériques) et des données inférées à partir des données volontaires et observées. La data science, littéralement « science des données », est la discipline qui étudie les informations et les méthodes permettant de les transformer en ressources « utiles ».

Les algorithmes et systèmes d'IA peuvent reposer, ou non, sur des traitements de données personnelles.

On distingue aujourd'hui deux grands types d'algorithmes :

- Les **algorithmes dits fermés**, qui sont des outils classiques : l'intégralité des instructions a été pensée et écrite par un être humain dans un code source. Ils brassent les seules données *ad hoc* préalablement déterminées, c'est-à-dire celles que le responsable de l'algorithme a identifiées en amont, et que la personne a volontairement et ponctuellement fournies pour un usage précis (par exemple Parcoursup, où la liste des données personnelles traitées par l'algorithme est établie dans un texte réglementaire) ;
- Les **algorithmes apprenants (apprentissage automatique ou *machine learning*)**, qui fonctionnent via un procédé par lequel les informations pertinentes sont tirées d'un ensemble de données d'entraînement, souvent dans le but de prédire ou d'estimer un résultat. L'humain indique uniquement à l'algorithme le résultat à atteindre à partir de données d'apprentissage qui peuvent lui être fournies. Le programme génère alors de lui-même les opérations lui permettant d'aboutir au résultat demandé.²¹
- Depuis une quinzaine d'années, se développe l'**apprentissage profond (ou *deep learning*)** : il s'agit d'un procédé qui relève de l'apprentissage automatique et qui

utilise des réseaux de neurones artificiels possédant plusieurs couches de neurones. Ces algorithmes possèdent de très nombreux paramètres et demandent un nombre très important de données afin d'être entraînés²², c'est-à-dire de construire un modèle à partir de données.

Les algorithmes apprenants constituent le fondement de ce que l'on nomme les systèmes d'IA. Selon le règlement sur l'IA adopté par l'Union européenne en 2024, un système d'IA est « *un système basé sur une machine qui est conçu pour fonctionner avec différents niveaux d'autonomie et qui peut faire preuve d'adaptabilité après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des données qu'il reçoit, comment générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer des environnements physiques ou virtuels* »²³.

Ces algorithmes posent ainsi des enjeux différents :

- L'utilisation d'algorithmes, simples ou apprenants, soulève des enjeux en termes de transparence, car les personnes qui les utilisent mais aussi celles vis-à-vis desquelles les algorithmes sont utilisés doivent pouvoir accéder à un certain niveau de compréhension du rôle de ces systèmes dans la décision prise et de leur logique interne ;
- Les systèmes apprenants, parce qu'ils se caractérisent par une capacité à élaborer eux-mêmes les règles leur permettant d'atteindre le résultat demandé, posent un enjeu spécifique en termes d'opacité. En effet, même les concepteurs de ces algorithmes « *perdent la capacité à comprendre la logique des résultats produits* »²⁴.

Ces enjeux seront détaillés dans la dernière partie de ce rapport.

Les algorithmes fermés et les systèmes d'IA, avec des exemples dans des secteurs variés de l'action administrative, sont ici examinés dans la mesure où ils servent de plus en plus à la prise de décisions administratives individuelles.

2· L'ALGORITHMISATION DE L'ADMINISTRATION

Dans la lignée de la **dématérialisation des procédures administratives**, que l'on peut définir comme l'essor des téléprocédures dans les services publics, les pouvoirs publics ont accéléré le déploiement des algorithmes et systèmes d'IA au sein de l'administration. L'essor du *big data*, notamment, autorise en effet un tel déploiement. Aujourd'hui, de nombreux services relevant de

l'administration s'appuient d'ores et déjà sur des algorithmes, à tel point que l'on peut parler de **l'algorithmisation de l'administration**²⁵ ou **d'action publique algorithmique**²⁶. On entend par là le recours croissant, par l'administration, à des algorithmes ou à des systèmes d'IA, visant à permettre d'automatiser, standardiser voire améliorer et/ou accélérer certaines procédures ou certaines étapes de procédures administratives.

FOCUS

DES STRUCTURES DÉDIÉES À LA TRANSFORMATION NUMÉRIQUE DE L'ÉTAT

Sur le plan administratif, l'algorithmisation et la numérisation de l'État ont été pilotées par des structures aux contours évolutifs, au cours des dernières années. Depuis 2019, cette action interministérielle est portée par deux grandes directions :

- La **direction interministérielle de la transformation publique (DITP)**, d'une part, qui anime et coordonne le programme de transformation publique et suit la bonne mise en œuvre des politiques prioritaires du gouvernement. Elle accompagne les ministères dans leurs chantiers de transformation, conçoit et déploie des projets interministériels clés pour la transformation de l'État ;
 - La **direction interministérielle du numérique (Dinum)**, d'autre part, qui a pour mission d'élaborer et de piloter la stratégie numérique de l'État. Elle compte parmi ses objectifs de développer l'exploitation effective des données « pour un État plus efficace dans son action et plus simple vis-à-vis des citoyens, des entreprises et des agents publics », via, notamment, le développement de traitements algorithmiques²⁷.
- > Au sein de la Dinum, le département **Etalab**, créé en 2011, coordonne notamment la conception et la mise en œuvre de la stratégie de l'État dans le domaine de la donnée, autour de 3 axes :

- Ouverture (*open data*) et partage des données : Etalab coordonne les actions des administrations de l'État et leur apporte son appui pour faciliter la diffusion et la réutilisation de leurs informations publiques ;
- Exploitation des données, intelligence artificielle et algorithmes publics : Etalab coordonne et promeut l'action de l'État et des organismes placés sous sa tutelle en matière d'inventaire, de gouvernance, de production, de circulation, d'exploitation et d'ouverture des données, et notamment des codes sources ;
- Innovation et ouverture de l'action publique : Etalab contribue à ouvrir l'administration et l'action publique sur la société, en mettant en œuvre les principes de « gouvernement ouvert » (transparence de l'action publique, consultation et concertation avec la société civile, participation citoyenne, innovation ouverte).

Etalab accompagne les administrations dans l'ouverture et l'usage responsable des algorithmes publics. À ce titre, il a, par exemple, publié en 2019, pour l'ensemble des administrations et organisations chargées d'une mission de service public qui conçoivent, développent et opèrent des traitements algorithmiques, un Guide des algorithmes publics²⁸, afin de les guider dans l'application du cadre juridique relatif aux algorithmes publics et notamment pour ce qui concerne la transparence requise.

Une catégorisation des différents types d'usages des algorithmes identifiés dans les administrations, a été effectuée par Etalab :

| USAGES | EXEMPLES |
|--|---|
| Attribuer des droits, calculer des montants selon des règles prédéfinies | Calcul des taxes et des impôts, attribution de prestations sociales, etc. |
| Réaliser un appariement entre une « offre » et une « demande » | Gestion de la mobilité des agents (RH), accès à l'enseignement supérieur (Parcoursup), attribution de places en crèches, etc. |
| Prédire une situation ou un risque en analysant des données | Prédire un risque de défaillance d'une entreprise (signaux faibles), cibler les contrôles de lutte contre la fraude fiscale, etc. |
| Aider à la décision des usagers | Aider les demandeurs d'emploi à cibler leurs candidatures spontanées (La Bonne Boîte), simuler le coût d'une embauche, etc. |

Les administrations, les collectivités territoriales, la justice, les services publics de façon générale **mobilisent d'ores et déjà et de plus en plus d'algorithmes et de systèmes d'IA**, seuls ou avec des instituts de recherche ou des entreprises privées prestataires, parfois à titre expérimental, pour analyser des masses de données et par exemple :

- procéder à l'anonymisation des décisions de justice²⁹,
- identifier des événements dans l'espace public à des fins « sécuritaires » via la vidéo surveillance augmentée³⁰,
- améliorer le diagnostic et la prévention en matière médicale³¹,
- personnaliser les apprentissages³²,
- lutter contre le gaspillage alimentaire dans les cantines des établissements publics³³,
- guider le déploiement de politiques publiques, simuler les conséquences économiques et sociales qu'un nouveau dispositif législatif est susceptible de produire³⁴,
- analyser des besoins territoriaux (gestion de l'eau et prévention/détection des fuites, ou encore mise en œuvre et optimisation de l'éclairage connecté).

Aujourd'hui, ce sont notamment les développements liés à l'utilisation de systèmes d'IA générative qui font parler d'eux. Selon la Cnil : « *Une intelligence artificielle générative est un système capable de créer du texte, des images ou d'autres contenus (musique, vidéo, voix, etc.) à partir d'une instruction d'un utilisateur humain. Ces systèmes peuvent produire des nouveaux contenus à partir de données d'entraînement.* »³⁵.

Alors que Chat GPT est aujourd'hui mondialement connu, plusieurs systèmes d'IA générative ont été développés dans le champ des services publics. Un système d'IA générative a par exemple été développé pour résumer les amendements déposés sur un projet ou une proposition de loi (« L'amendement »). Pour ce qui concerne les relations entre les usagers et l'administration, c'est « Albert », autre exemple de système d'IA générative, qui est développé depuis juin 2023.

EXEMPLE**LE SYSTÈME D'IA GÉNÉRATIVE « ALBERT »³⁶**

Albert est le nom donné à un système d'IA **généraliste**, développé par l'État pour « **accélérer les formalités administratives** et apporter des réponses sûres, claires et efficaces aux utilisateurs des **services publics** ».

Cet assistant conversationnel est développé par la Dinum à partir de briques existantes (notamment les fiches pratiques de service-public.fr.). Tchap Albert est, à ce stade, réservée aux agents de l'administration, pour les aider à mieux répondre aux usagers, rechercher des informations sur la réglementation et même rédiger des synthèses. Il commence à être déployé dans les espaces France services.

Il est d'ores et déjà prévu les autres usages suivants :

- **sous-titrage automatique de cours magistraux** ;
- **retranscription d'audiences judiciaires** ;
- **rédaction de dépôts de plainte ou de comptes rendus médicaux**.

Le recours par l'administration à ces procédés requiert une vigilance particulière, pour deux raisons³⁷ :

- **Les algorithmes sont utilisés pour faire appliquer des dispositions normatives** (loi, décret...) ;
- Contrairement à un service privé, auquel on peut – dans une certaine mesure – décider de ne pas avoir recours, **les usagers des services publics n'ont pas d'autre choix qu'y recourir**.

Ceci étant, si le déploiement d'une action publique « algorithmique » soulève plusieurs questions, certains avantages peuvent être associés à l'utilisation de systèmes automatisés par l'administration dans ses relations avec les usagers³⁸ :

- Elle peut permettre un processus décisionnel plus efficace et mieux informé par le traitement d'une multitude de données ;
- Elle peut impliquer d'édicter et de formaliser des critères qui permettent de retracer la logique des décisions prises, là où ces mêmes décisions pouvaient auparavant être prises de façon opaque (par exemple, l'attribution de places en crèche) ;
- De ce fait, elle peut contribuer à limiter l'arbitraire et permettre d'atteindre une égalité de traitement, les agents humains n'étant pas toujours neutres et leurs décisions pouvant s'avérer discriminatoires.

Ces avantages, pour se concrétiser, impliquent une appropriation et une utilisation des algorithmes et systèmes d'IA qui soit responsable et conforme au droit, au risque, sinon, de problèmes critiques.

3- L'OUVERTURE À LA PRISE DE DÉCISIONS ADMINISTRATIVES INDIVIDUELLES ENTIÈREMENT AUTOMATISÉES

Depuis la promulgation de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, **la prise de décision individuelle entièrement automatisée est interdite dans le champ de l'action administrative**. Cette volonté du législateur avait été mise au jour dès le stade du rapport fait au nom de la Commission des Lois constitutionnelles du Sénat sur le projet de loi, qui insistait, au regard de la place grandissante des ordinateurs, sur la nécessité de « *faire qu'en aucune manière ce mode de jugement ne supprime les moyens traditionnels et introduise l'automatique là où la nuance, pour ne pas dire la délicatesse, sont souvent de mise* »³⁹.

Malgré les différentes modifications apportées à cette loi, ce principe d'interdiction subsiste encore aujourd'hui explicitement **pour les décisions de justice** qui impliquent « *une appréciation sur le comportement d'une personne* »⁴⁰.

Il en est allé différemment dans le champ de l'action administrative.

S'il était acquis que l'administration se fondait déjà sur des algorithmes pour aider à la prise de décision (pour le calcul des droits, l'affectation de personnels ou d'élèves par exemple), **la volonté du Gouvernement en 2018 a été d'« ouvrir plus largement la possibilité pour l'administration de recourir à des décisions automatisées (prises sur le fondement d'un algorithme) »**⁴¹. Cela a été fait via la loi relative à la protection des données personnelles⁴².

Dans le même sens, il était souligné dans l'étude d'impact qu'« *une intervention humaine ne semble pas toujours nécessaire pour assurer les droits des personnes concernées et des usagers des services publics, comme par exemple pour le calcul de l'impôt sur le revenu. Le maintien de cette interdiction absolue ne permet pas de répondre aux évolutions de l'activité administrative qui a de plus en plus recours à des traitements algorithmiques, notamment pour les décisions de masse que la réglementation encadre précisément et dont l'édition rapide permet la bonne délivrance du service public* »⁴³. Deux exemples relevant du champ des finances publiques étaient mis en avant : le calcul des prélèvements et des droits d'une part, la lutte contre la fraude d'autre part. L'idée centrale était que l'efficacité administrative requerrait de pouvoir passer par ce type de décisions moyennant certaines conditions.

Le principe d'interdiction a donc été assorti d'**exceptions qui ouvrent, de fait, de larges possibilités de recourir aux décisions administratives individuelles entièrement automatisées**. Ainsi, dans le droit en vigueur, s'agissant de l'action administrative :

- Il existe toujours un principe d'interdiction pour les décisions entièrement automatisées qui « *produisent des effets juridiques pour les personnes* » ou « *qui les affectent de manière significative* »⁴⁴, un champ au contenu à la fois très large et à l'identification potentiellement complexe ;
- Dans le sillage ouvert par le RGPD⁴⁵ au niveau de l'Union européenne, le législateur a décidé⁴⁶ que **les décisions administratives individuelles respectant certaines conditions pouvaient être prises « sur le seul fondement d'un traitement automatisé de données à caractère personnel »**.

A noter qu'on s'intéresse ici aux décisions administratives individuelles qui ne relèvent pas du champ pénal au sens de la directive n° 2016/680 du 27 avril 2016, dite directive « Police-Justice »⁴⁷, laquelle prévoit un régime d'exception par rapport au RGPD.

Les conditions encadrant la prise de décisions administratives individuelles entièrement automatisées, qui constituent **des garanties pour les usagers du service public**, sont les suivantes :

1. **L'information de la personne concernée** par cette décision, du fait que cette décision a été prise sur le fondement d'un algorithme et les éléments principaux de compréhension de celui-ci ;
2. **La maîtrise du traitement algorithmique** par l'organisme public qui l'utilise ;
3. La possibilité, pour la personne concernée par cette décision, d'**obtenir une intervention humaine en cas de recours** contre la décision entièrement automatisée ;
4. L'impossibilité que la décision entièrement automatisée soit prise **à partir de « données sensibles »** au sens de la loi « Informatique et Libertés ».

Le Conseil constitutionnel a validé ces dispositions en éclairant les conditions prévues par le législateur, notamment en relevant que « *ne peuvent être utilisés, comme fondement exclusif d'une décision administrative individuelle, des algorithmes susceptibles de réviser eux-mêmes les règles qu'ils appliquent, sans le contrôle et la validation du responsable du traitement* »⁴⁸. C'est donc au regard des garanties apportées que l'exception ouverte pour l'action administrative a été jugée conforme aux exigences qui sous-tendent notre cadre juridique.

Plus de cinq ans après la loi, il apparaît nécessaire de dresser un bilan de l'équilibre global du dispositif et de la mise en œuvre de ces garanties.

FOCUS

LES DONNÉES PERSONNELLES DITES SENSIBLES ET LES RISQUES DE DISCRIMINATIONS AUTOMATISÉES

Le risque d'automatisation des discriminations induit par l'utilisation de systèmes algorithmiques ou d'IA a été largement démontré⁴⁹. Et des exemples bien réels montrent que ce risque a malheureusement pu se concrétiser, y compris dans le secteur public.

> Parmi eux, on peut citer le cas de l'algorithme développé en Autriche par l'agence nationale pour l'emploi « AMS » en 2019, dont le but était d'attribuer un « score d'employabilité » aux chômeurs. Des régressions statistiques ont été effectuées pour déterminer les facteurs les plus susceptibles de prédire les chances d'un individu de trouver un emploi et le type d'accompagnement à proposer compte tenu des ressources limitées de l'agence.

Selon l'enquête menée par l'association AlgorithmWatch⁵⁰, le système déployé attribuait automatiquement un score à chaque demandeur d'emploi sur la base de plusieurs caractéristiques. En fonction du score obtenu, les demandeurs d'emploi étaient répertoriés dans 3 groupes :

- Le groupe A pour les personnes n'ayant pas besoin d'aide pour trouver un nouvel emploi ;
- Le groupe B pour les personnes pouvant bénéficier d'une reconversion professionnelle et d'un accompagnement ;
- Le groupe C pour les personnes jugées inaptes au travail, recevant de ce fait moins d'aide de l'AMS et étant orientées vers d'autres institutions.

L'analyse des résultats du système, une fois rendue publique, a fait l'objet de vives critiques quant aux biais discriminatoires et au respect du droit de la non-discrimination : les femmes chômeuses se voyaient attribuer un poids négatif, tout comme les personnes en situation de handicap et les personnes de plus de 30 ans. Les femmes ayant des enfants faisaient aussi l'objet d'une évaluation négative, ce qui n'était pas le cas des hommes ayant des

enfants. L'algorithme avalisait ce faisant les discriminations systémiques et les inégalités d'accès à l'emploi du marché. Le système, déclaré illégal dans un premier temps par l'homologue autrichien de la Cnil, a été admis en appel ; cependant l'algorithme de l'AMS n'est plus utilisé, à la suite des polémiques suscitées.

Parmi les garanties adoptées en France au moment de l'ouverture à la prise de décisions administratives individuelles entièrement automatisées, **l'interdiction d'en prendre sur le fondement de traitement de données personnelles dites « sensibles » a été posée comme permettant en partie de lutter contre d'éventuelles discriminations**⁵¹.

L'article 9 du RGPD définit comme suit un tel traitement : « *traitement des données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique* ».

Or, si cette définition paraît claire, son interprétation récente par des tribunaux français et européen suscite des interrogations. Notamment, il apparaît que :

- La Cour de justice de l'Union européenne (CJUE) tend, dans plusieurs décisions récentes, à considérer qu'une appréhension large de cette catégorie doit être retenue, y compris quand celle-ci relève d'une approche indirecte⁵² ;
- Le statut de la donnée personnelle qui révèle la présence d'un handicap⁵³, et est donc potentiellement assimilable à une donnée personnelle concernant la santé peut s'en trouver transformé.

Des interprétations extensives de la notion de donnée sensible, qui relève en premier lieu de la compétence de la Cnil sous le contrôle du Conseil d'État, pourraient venir restreindre les possibilités de l'administration de prendre des décisions entièrement automatisées.

Par ailleurs, la définition des données sensibles au sens du RGPD ne recoupe pas exactement celle des caractéristiques protégées par le droit de la non-discrimination⁵⁴. En effet, la situation de famille (parent isolé notamment), la vulnérabilité résultant de la situation économique, le sexe, voire le lieu de résidence sont des motifs protégés en droit de la non-discrimination, mais non pas des données considérées comme sensibles au sens de la réglementation applicable en matière de protection des données.

Enfin, il faut avoir à l'esprit que les algorithmes apprenants peuvent intégrer lors de leur développement des indicateurs, des variables équivalentes (appelées *proxies*). Ces *proxies* peuvent tout à fait se substituer alors à un critère de discrimination : un indicateur sur les personnes ayant acheté des tampons permet aussi de mesurer les femmes (d'une certaine tranche d'âge) et de cibler ces dernières le cas échéant. Cette situation d'« encodage redondant », c'est-à-dire l'appartenance à une catégorie protégée encodée dans d'autres données, n'est pas nécessairement intentionnelle mais découle du programme apprenant.

FOCUS

LA NOTION DE « DÉCISION »

La nécessité de retenir une interprétation large de la notion de « décision ».

Dans un arrêt récent⁵⁵, la CJUE s'est intéressée à l'attribution par une société allemande d'un score de solvabilité à une personne physique, *scoring* ensuite communiqué à une banque. La cour a considéré que dans le cas où la valeur de probabilité établie par la société joue un rôle déterminant dans l'octroi d'un crédit par la banque, ce *scoring* doit être qualifié en soi de décision produisant à l'égard d'une personne concernée « *des effets juridiques la concernant ou l'affectant de manière significative de façon similaire* ».

La Cour souligne que cette interprétation, et surtout la portée large de la notion de « décision », renforce la protection effective visée par le RGPD. En revanche, une interprétation restrictive, selon laquelle

l'établissement de la valeur de probabilité doit seulement être considéré comme un acte préparatoire, seul l'acte adopté par la tierce partie pouvant, le cas échéant, être qualifié de « décision », entraînerait une lacune dans la protection juridique. En effet, dans cette hypothèse, l'établissement du *scoring* échapperait aux exigences du RGPD qui interdit par principe la prise de décision entièrement automatisée quand cette décision emporte des effets juridiques ou significatifs pour la personne concernée, tout en admettant des exceptions. En effet, cette procédure repose sur un traitement automatisé de données personnelles et produit des effets affectant de manière significative la personne concernée dans la mesure où l'action de la tierce partie, à laquelle la valeur de probabilité est transmise, est guidée de manière déterminante par celle-ci.

Sur la base de cet arrêt, il conviendra de déterminer si le périmètre des décisions administratives individuelles, visées au 2° de l'article 47 de la loi « Informatique et Libertés », doit être clarifié.

D'ores et déjà, la question se pose s'agissant des « scores de risque » attribués par les caisses d'allocations familiales à leurs allocataires⁵⁶, au regard de l'incidence de ce score sur la mise en œuvre de contrôles de lutte contre l'erreur et la fraude.

La distinction aujourd'hui retenue dans le droit en vigueur s'intéresse à « la décision » d'une façon qui peut paraître insuffisante car ne couvrant pas forcément le processus d'élaboration de celle-ci. En effet, la décision prise par l'administration peut être la dernière étape d'un processus qui en fait intervenir plusieurs. Que se passe-t-il si l'une de ces étapes, entièrement automatisée, joue un rôle déterminant dans la décision ?

Le Conseil constitutionnel, en se prononçant sur les dispositions de la loi de 2018 qui a ouvert la porte aux décisions administratives individuelles entièrement automatisées, a considéré que l'administration était autorisée à « *procéder à l'appréciation individuelle de la situation de l'intéressé, par le seul truchement d'un algorithme, en fonction des règles et critères définis à l'avance par le responsable du traitement* »⁵⁷. C'est l'enjeu de l'analyse relative à l'intervention humaine.

PARTIE II

LA MAÎTRISE ET L'INTERVENTION HUMAINE : DES GARANTIES, MAIS CONCRÈTEMENT ?

En l'état du droit, l'un des enjeux majeurs, pour ce qui concerne les décisions administratives individuelles, est d'établir si elles sont entièrement ou seulement partiellement automatisées, ce qui revient à interroger la substance de l'intervention humaine.

Il est auparavant nécessaire de replacer cette question dans un cadre plus large : celui du contrôle humain des systèmes automatisés qui permettent de prendre des décisions individuelles mais aussi de guider des politiques publiques (par exemple, pour anticiper la propagation d'une épidémie et identifier les territoires les plus concernés).

1· LES DIFFÉRENTS NIVEAUX DE L'INTERVENTION HUMAINE

Le concept de « supervision humaine » renvoie à la nécessité pour les humains de « garder la main » sur les systèmes utilisés⁵⁸ afin que leur fonctionnement ne devienne pas complètement autonome et au fait que le responsable -ici l'administration- ne saurait invoquer l'utilisation de l'algorithme ou du système d'IA utilisé pour prendre une décision administrative individuelle (que celle-ci soit complètement ou partiellement automatisée) si celui-ci était mis en cause dans le cadre d'une contestation.

Dans les textes en vigueur, on retrouve diverses formulations pour désigner ce rôle de supervision. Les textes mobilisent différentes notions et précisent des obligations différentes. Pour exemple :

- La notion de « maîtrise » par le responsable du traitement figure à l'article 47 de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés ;
- La notion d' « intervention humaine » à l'occasion d'une décision individuelle automatisée ou d'un profilage est inscrite à l'article 22 du RGPD ;
- La notion de « contrôle humain » a été introduite à l'article 14 du règlement de l'UE sur l'IA.

Ces vocables sont utilisés dans le présent rapport de façon substitutive, pour parler de l'intervention ou du contrôle humain qui peut et doit intervenir à différentes étapes.

Le contrôle humain de ces systèmes s'effectue en effet principalement à deux niveaux :

- Le contrôle « système » qui concerne le fonctionnement du système dans son ensemble ;
- Le contrôle « individuel » qui concerne une décision individuelle.



LES CONTRÔLES « SYSTÈME »

Trois phases distinctes de contrôle peuvent être distinguées : pendant la phase de conception (phase 1), pendant l'exploitation (phase 2), pendant les tests et audits ultérieurs (phase 3). Il s'agit de trois types de contrôle qui interviennent lors de phases que l'on peut en théorie distinguer mais qui peuvent se chevaucher (les audits pouvant avoir lieu pendant l'exploitation).

L'intervention humaine « *ab initio* », dans l'édiction des règles par l'algorithme ou le système d'IA (phase 1).

Quand l'algorithme est mis en œuvre par une administration, les choix effectués doivent être conformes aux textes qui s'appliquent dans le domaine concerné - outre le cadre juridique applicable en général aux algorithmes. Par exemple, l'Agence de la biomédecine utilise un système algorithmique pour établir l'ordre de proposition des greffons cardiaques, appelé « score cœur ».

Concrètement, à chaque fois qu'un greffon cardiaque est disponible, le système calcule un score pour l'ensemble des patients sur la liste d'attente. Les règles de cet algorithme doivent notamment prendre en compte, conformément aux textes sectoriels, la priorité pour les moins de 18 ans « *selon des modalités définies par l'Agence de la biomédecine et tenant compte de l'urgence à être greffé* »⁵⁹.

Le service public qui déploie l'algorithme ou le système d'IA est responsable des règles et critères de fonctionnement et il doit pouvoir en rendre compte. En outre, les règles définissant les principaux traitements algorithmiques utilisés dans l'accomplissement des missions, lorsque ces traitements fondent des décisions individuelles (en partie ou entièrement), doivent être publiées⁶⁰.

EXEMPLE**L'ALGORITHME LOCAL DÉVELOPPÉ ET UTILISÉ PAR SCIENCES PO BORDEAUX DANS LE CADRE DE LA PROCÉDURE PARCOURSUP**

La procédure Parcoursup constitue la procédure principale d'accès aux établissements de l'enseignement supérieur. Elle sera présentée en détail ci-après. Schématiquement, la procédure comprend trois phases distinctes :

- Les candidats enregistrent leurs vœux sur la plateforme nationale gérée par le ministère de l'enseignement supérieur et de la recherche ;
- Les établissements d'enseignement supérieur classent les dossiers de candidature qu'ils reçoivent via la plateforme nationale en utilisant bien souvent un algorithme local pour évaluer ces candidatures (l'outil proposé par le ministère, ou un autre) puis lui renvoie une liste d'appel ;
- L'algorithme de la plateforme nationale fait le lien entre les vœux de chaque candidat et le classement des dossiers de candidature réalisé par les établissements d'enseignement supérieur puis transmet les propositions faites à chaque candidat.

Sciences Po Bordeaux indique utiliser un algorithme de sélection des dossiers des candidats qui n'est pas l'outil d'aide à la décision (OAD) proposé par le ministère⁶¹. Dans ce cadre, l'établissement ne peut pas sélectionner d'autres données personnelles que celles prévues par les textes⁶² ; en revanche, il peut mobiliser des critères qu'il juge pertinents.

En l'occurrence, le choix a été fait de se fonder, dans une certaine mesure, non sur les seules notes (critère retenu pour l'OAD proposé par le ministère) mais sur l'écart entre les notes de l'élève et la moyenne de la classe. L'outil, qui permet d'évaluer plus finement le niveau des élèves, est ainsi défavorable aux élèves des lycées qui pratiquent une forme d'inflation des notes et à ceux dont l'homogénéité du niveau

des élèves se traduit par un manque de dispersion des notes au sein de la classe⁶³. Cet algorithme de sélection des dossiers est aussi paramétré pour favoriser les élèves boursiers (au-delà du quota obligatoire prévu par la loi) et ceux qui sont issus de lycées participant aux « cordées de la réussite ».

Le concepteur de l'algorithme, le professeur Vincent Tiberj, revendique « un choix sociologique » et explique que la conséquence des choix de cet algorithme dit « maison » est « bien de favoriser les lycées qui acceptent une diversité au sein de leurs classes ».

- > Les résultats sont frappants. En 2023, le profil des élèves considérés comme admissibles en utilisant l'algorithme dit « maison » et celui des élèves considérés comme admissibles en utilisant l'OAD proposé par le ministère variait sensiblement :
- Les élèves venant d'une cordée de la réussite passent de 19 % des admissibles dans le premier groupe à 3 % dans le second ;
 - Les élèves boursiers passent de 12 % à 5 % ;
 - Les élèves originaires des DOM-TOM passent de 8,5 % à 3,5 %.

Ce cas illustre le fait qu'il y a derrière cet algorithme comme pour d'autres un choix politique (et non pas seulement technique) qui est fait par des humains, des décisions qui sont prises, voire des effets qui sont anticipés.

Le paramétrage de ces algorithmes doit respecter des textes légaux (par exemple, en termes de respect de la liste des données personnelles traitées) et traduire les objectifs politiques que l'administration concernée se donne. Parce qu'il existe un risque d'écart entre le cadre juridique et administratif et le code, qui ne reposent pas sur le même langage⁶⁴, cela implique de développer une réflexion partagée sur ces outils et de disposer notamment de



compétences techniques (savoir coder) et de temps, que les administrations ne peuvent pas nécessairement assurer. On le constate notamment dans les établissements d'enseignement supérieur, qui pour la très grande majorité, se contentent dans le cadre de la procédure Parcoursup, d'utiliser l'outil d'aide à la décision (OAD) proposé par le ministère.

Dans les systèmes d'apprentissage automatique, c'est l'algorithme qui définit lui-même les règles. L'humain peut néanmoins intervenir aux étapes suivantes (liste non exhaustive) :

- Définition de l'objectif à atteindre ;
- Sélection des données avec lesquelles le système va apprendre (« données d'apprentissage » ou « jeu d'apprentissage ») ;
- Étiquetage ou annotation de ces données (cette étape peut aussi être automatisée : elle consiste à décrire la donnée) ;
- Choix des critères retenus pour évaluer les résultats du système (on parle aussi de « métriques »), par exemple, le taux de faux positifs⁶⁵ et de faux négatifs⁶⁶.

Par exemple, en matière de gestion des déchets, des municipalités peuvent concevoir des systèmes d'IA pour mesurer le taux de remplissage des poubelles et adapter en conséquence les interventions des agents en charge.

L'intervention humaine au sens de la supervision du système pendant son exploitation et lors des tests et audits ultérieurs (phases 2 et 3).

Pendant ces phases, il s'agit :

- de surveiller le fonctionnement du système, l'arrêter à tout moment en cas d'anomalie ;
- de détecter les erreurs et les biais, en sus du travail déjà effectué en ce sens pendant la conception.

Ces étapes sont particulièrement importantes lors des mises à jour des algorithmes d'apprentissage automatique qui présentent la particularité de réviser eux-mêmes leurs propres règles et donc d'être évolutifs. Ils contiennent donc un risque accru d'erreur d'une part et de discrimination d'autre part, du fait des corrélations qu'ils créent.

C'est la raison pour laquelle le règlement sur l'IA de l'UE a créé des obligations en la matière.

FOCUS**DISPOSITIONS DU RÈGLEMENT SUR L'IA DE L'UE POUR LES SYSTÈMES D'IA « À HAUT RISQUE »**

Le règlement sur l'IA de l'UE soumet les systèmes d'IA considérés comme à haut risque⁶⁷ (listés notamment dans l'annexe III du règlement) au principe du « *contrôle humain* » (en anglais, *human oversight*). L'objectif est de réduire au minimum les risques pour la santé, la sécurité ou les droits fondamentaux d'un système d'IA déployé par l'utilisateur⁶⁸ de façon conforme à sa destination – c'est-à-dire telle que précisée par le fournisseur – ou dans des conditions de mauvaise utilisation « raisonnablement prévisible » selon les termes dudit règlement.

Certains des systèmes d'IA identifiés par le règlement sur l'IA de l'UE comme « à haut risque » concernent le fonctionnement des services publics, puisqu'on y trouve notamment :

- *Dans le champ de l'accès et du droit aux services privés essentiels et aux services publics et prestations sociales essentiels, les systèmes d'IA destinés à :*
 - > Être utilisés par les autorités publiques ou en leur nom pour évaluer l'éligibilité des personnes physiques aux prestations et services d'aide sociale essentiels, y compris les services de soins de santé, ainsi que pour octroyer, réduire, révoquer ou récupérer ces prestations et services ;
 - > Être utilisés pour l'évaluation des risques et la tarification applicable aux personnes physiques en matière d'assurance-vie et d'assurance maladie ;
 - > Évaluer et hiérarchiser les appels d'urgence émanant de personnes physiques ou à être utilisés pour envoyer ou établir des priorités dans l'envoi des services d'intervention d'urgence, y compris par la police, les pompiers et l'assistance médicale, ainsi que pour les systèmes de tri des patients admis dans les services de santé d'urgence.

- *Dans le champ de l'éducation et de la formation professionnelle, les systèmes d'IA destinés à être utilisés :*

- > Pour déterminer l'accès, l'admission ou l'affectation de personnes physiques à des établissements d'enseignement et de formation professionnelle ;
- > Pour évaluer les acquis d'apprentissage, y compris lorsque ceux-ci sont utilisés pour orienter le processus d'apprentissage de personnes physiques dans les établissements d'enseignement et de formation professionnelle ;
- > Pour évaluer le niveau d'enseignement approprié qu'une personne recevra ou sera en mesure d'atteindre, dans le contexte ou au sein d'établissements d'enseignement et de formation professionnelle ;
- > Pour surveiller et détecter des comportements interdits chez les étudiants lors d'examens.

Précisément, l'article 14 du règlement relatif au contrôle humain prévoit que tant la conception que le développement de ces systèmes à haut risque permettent « *notamment au moyen d'interfaces homme-machines appropriées, un contrôle effectif par des personnes physiques pendant leur période d'utilisation* ». Il est précisé que les mesures mises en œuvre pour ce contrôle, qui doivent être proportionnées aux risques, sont assurées *via* :

- Des mesures identifiées et, lorsque cela est techniquement possible, intégrées par le fournisseur dans le système d'IA à haut risque avant la mise en service de ce dernier ;
- Des mesures identifiées par le fournisseur du système avant la mise en service de ce dernier et qui se prêtent à une mise en œuvre par le déployeur.

Concrètement, selon le règlement, les personnes physiques chargées d'effectuer un contrôle humain pour le compte de l'organisme qui déploie le système d'IA doivent avoir la possibilité de :

- Comprendre correctement les capacités et les limites pertinentes du système, et être en mesure de surveiller correctement son fonctionnement, y compris en vue de détecter et de traiter les anomalies, les dysfonctionnements et les performances inattendues ;
- Avoir conscience d'une éventuelle tendance à se fier automatiquement ou excessivement aux résultats issus du système (biais d'automatisation), en particulier pour les systèmes utilisés pour fournir des informations ou des recommandations concernant les décisions à prendre par les personnes physiques ;
- Interpréter correctement les résultats du système, compte-tenu des outils et méthodes d'interprétation disponibles ;
- Décider, dans une situation particulière, de ne pas utiliser le système ou encore d'ignorer, de remplacer voire d'inverser la « sortie » du système ;
- Intervenir dans le fonctionnement du système ou l'interrompre au moyen d'un bouton d'arrêt ou d'une procédure similaire.

À l'heure où ce rapport est rédigé, des normes harmonisées sur lesquelles vont reposer de façon précise et concrète les exigences posées par le règlement sur l'IA, et notamment les obligations prévues à son article 14, sont en cours d'élaboration. Le comité technique JTC-21⁶⁹ mis en place à la demande de la Commission européenne est l'organisme responsable de ce processus. Schématiquement, les systèmes d'IA qui respecteront ces futures normes seront présumés conformes aux obligations posées par le règlement sur l'IA.

Si la Défenseure des droits se félicite que l'impératif de contrôle humain du système pendant son utilisation soit organisé à l'article 14 du règlement, il conviendra de rester attentif au contenu de la future norme harmonisée qui doit venir préciser les exigences du texte.

En outre, l'article 14 du règlement sur l'IA de l'UE ne vise donc que les systèmes d'IA considérés comme à haut risque, ceux considérés comme à risque limité et à risque minimum n'étant pas concernés. Et il faut avoir à l'esprit que le règlement a prévu un système de filtre⁷⁰ qui permet de déclasser un système d'IA qui ferait sinon partie de la catégorie à haut risque au regard de son domaine d'intervention. Un système peut être extrait de la catégorie à haut risque aux conditions (alternatives) suivantes :

1. Si le système d'IA est destiné à accomplir une tâche procédurale étroite ;
2. Si le système d'IA est destiné à améliorer le résultat d'une activité humaine préalablement réalisée ;
3. Si le système d'IA est destiné à détecter les constantes en matière de prise de décision ou les écarts par rapport aux constantes habituelles antérieures et n'est pas destiné à se substituer à l'évaluation humaine préalablement réalisée, ni à influencer celle-ci, sans examen humain approprié ;
4. Si le système d'IA est destiné à exécuter une tâche préparatoire en vue d'une évaluation pertinente aux fins des cas d'utilisation visés à l'annexe III.

Par exception, un système d'IA sera toujours considéré comme étant à haut risque lorsqu'il effectue un profilage⁷¹ de personnes physiques.

Sur ce point, le règlement sur l'IA de l'UE renvoie au RGPD qui le définit comme suit (article 4.4) : « *toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique* ». Comme l'indique la Cnil « *Un traitement de profilage repose sur l'établissement d'un profil individualisé, concernant une personne en particulier : il vise à évaluer certains de ses aspects personnels, en vue d'émettre un jugement ou de tirer des conclusions sur elle (...). Tout traitement de profilage doit faire l'objet d'une attention particulière, car il soulève par nature des risques importants pour les droits et libertés des personnes* ». L'autorité donne l'exemple suivant : « *Le profilage est au cœur du secteur des assurances, notamment lorsqu'il implique l'évaluation des caractéristiques du risque assurantiel pour en déterminer la fréquence, le coût moyen, le coût maximum du sinistre potentiel, la tarification et vérifier l'assurabilité du risque. Il n'est donc qu'un moyen, non une finalité* »⁷².

À cet égard, la Défenseure des droits s'inquiète de la possibilité, pour les systèmes d'IA déployés par l'administration, les collectivités et autres organismes chargés d'une mission de service public dans leurs relations vis-à-vis des usagers (de façon à la fois générale, pour guider les politiques publiques mais aussi individuelle, pour la prise de décisions), de sortir de la catégorie « à haut risque » par le biais de ce système de filtre. Cela eu égard d'une part aux risques identifiés et d'autre part aux garanties que le règlement prévoit pour cette catégorie.

Le règlement sur l'IA de l'UE prévoit en outre d'autres obligations pour les systèmes d'IA à haut risque et notamment :

- Un système de gestion des risques (article 9), relatif notamment à l'identification et l'analyse des risques -connus et raisonnablement prévisibles, mais aussi susceptibles d'apparaître- que le système d'IA peut poser pour, entre autres, les droits fondamentaux, lorsque ce système est utilisé conformément à sa destination ainsi que l'adoption de mesures appropriées et ciblées pour répondre aux risques identifiés ;
- Des pratiques en matière de gouvernance et de gestion des données appropriées à la destination du système d'IA pour les jeux de données d'entraînement, de validation et de test, quand ce système contient des techniques qui impliquent l'entraînement de modèles d'IA au moyen de données (article 10) ;
- La tenue d'une documentation technique établie de manière à démontrer que le système d'IA satisfait aux exigences prévues (article 11) ;
- Une conception et un développement des systèmes d'IA tels que le fonctionnement de ces systèmes est suffisamment transparent pour permettre aux organismes qui les déploient d'interpréter les sorties d'un système et de les utiliser de manière appropriée, grâce notamment à une notice d'utilisation (article 13) ;
- Un niveau approprié d'exactitude, de robustesse et de cybersécurité, et un fonctionnement de façon constante à cet égard tout au long du cycle de vie de ces systèmes (article 15).

LES CONTRÔLES « INDIVIDUELS »

L'intervention humaine au moment de prendre, sur la base du résultat d'algorithmes ou de systèmes d'IA, la décision administrative individuelle.

Aujourd'hui, le droit en vigueur distingue deux types de décisions administratives individuelles selon leur degré d'automatisation ou d'automatisme⁷³ :

- D'une part, **les décisions administratives individuelles partiellement automatisées** : c'est-à-dire des décisions dans lesquelles un algorithme ou un système d'IA est intervenu à un moment donné, pour une finalité précise, mais où le résultat de ces systèmes n'a été qu'un élément dans la prise de décision, qui a pris en compte d'autres dimensions. L'intervention humaine est notamment requise lorsque le traitement de données personnelles mobilisé contient des données sensibles au sens du RGPD, d'où l'importance que le périmètre de ce qui relève ou non de cette catégorie soit clair.
 - > Par exemple, pour déterminer le montant attribué de la prestation de compensation du handicap (PCH), gérée par un département⁷⁴.
- D'autre part, **les décisions administratives individuelles entièrement automatisées** : c'est-à-dire les décisions qui sont prises par l'administration mais sans autre considération que le seul résultat de l'algorithme ou du système d'IA, qui à lui seul constitue la décision. Il n'y a donc pas, ici, d'intervention humaine au moment de la prise de décision individuelle. Pour autant, la décision reste bien juridiquement celle de l'administration ou de la collectivité.
 - > Par exemple, le calcul de la somme à payer au titre de l'impôt sur le revenu.

Le critère de répartition entre ces deux catégories de décision est **l'intervention humaine. Pour être qualifiée comme telle**, elle doit être consistante et avoir une influence réelle sur le résultat.

L'intervention humaine obligatoire en cas de recours administratif contre la décision individuelle.

En cas de recours administratif contre une décision individuelle (au sens des articles L. 410-1 et suivants du CRPA), une intervention humaine est requise⁷⁵. Ce recours appelle une attention particulière de la part de l'administration.

Ce type d'intervention ne concerne donc pas un traitement « de masse » et, au regard des garanties déjà prévues au titre Ier du livre IV du CRPA, n'appelle pas à ce stade d'observations particulières.

2· LA SUBSTANCE DE L'INTERVENTION HUMAINE AU MOMENT DE LA PRISE DE DÉCISION INDIVIDUELLE – RECOMMANDATIONS POUR LA GARANTIR

Pour caractériser une décision administrative individuelle comme seulement partiellement automatisée, l'intervention humaine ne peut se limiter à un « *simple geste symbolique* »⁷⁶. Cette intervention peut en particulier se traduire par l'exigence d'une étape supplémentaire : l'agent effectue une action positive, concrète et significative à partir ou à côté du résultat généré par l'algorithme. L'agent peut par exemple se voir confier la même tâche que l'algorithme, et doit ensuite, en cas de résultat différent (ex : calcul, identification) arbitrer entre les deux résultats celui qu'il convient de retenir.

Ainsi, le fait qu'un agent applique de façon systématique, par exemple en appuyant sur un bouton « valider », les résultats générés par l'algorithme ou le système d'IA, sans opération intellectuelle de vérification, ne saurait être considéré comme une intervention humaine de nature à faire de cette décision une décision partiellement automatisée. Il y a près de dix ans, le Conseil d'État insistait déjà sur la nécessité d'éviter que « *les systèmes présentés comme relevant de 'l'aide à la décision' soient en réalité presque toujours suivis et commandent la décision, l'intervention humaine n'étant alors qu'apparente* »⁷⁷. Une marge de manœuvre significative du côté des agents est donc requise pour considérer que le traitement automatisé n'a qu'une place partielle dans la prise de décision.

L'examen des délibérations de la Cnil fait apparaître des éléments de caractérisation de l'intervention humaine qui ont trait à l'existence d'étapes supplémentaires de la décision. Par exemple :

- « *l'examen individuel, permettant notamment de lever les cas d'homonymie, après collecte, le cas échéant, d'informations complémentaires (...)* »⁷⁸ ;
 - ou encore le fait que certains dispositifs de télésurveillance pour les examens en ligne ont pour « *seul rôle (...) d'attirer l'attention d'un surveillant sur une situation potentiellement anormale* », la « *vérification humaine* » devant systématiquement avoir lieu avant de confirmer la suspicion de fraude⁷⁹ ;
- > Saisi en 2020 de dispositions autorisant l'administration à recourir à des moyens informatisés et automatisés pour collecter et exploiter les contenus accessibles sur Internet afin de lutter contre la fraude fiscale, le Conseil constitutionnel a considéré qu'il n'y avait là aucune décision entièrement automatisée et validé le nouveau dispositif de contrôle pour le recouvrement de l'impôt. Les opérations de « *corroboration* » et d'« *enrichissement* » par les services compétents permettent en effet, selon le Conseil constitutionnel, de caractériser « *une appréciation individuelle de la situation de la personne par l'administration, qui ne peut alors se fonder exclusivement sur les résultats du traitement automatisé* »⁸⁰.

Il ressort de ces éléments que deux caractéristiques doivent toujours pouvoir être constatées : la personne qui intervient doit avoir pour cela l'autorité (que l'on pourrait définir comme le fait que l'organisation interne reconnaisse à cette personne la qualité pour agir) et la compétence (que l'on pourrait définir comme le fait d'avoir les moyens intellectuels et pratiques pour pouvoir exercer ce contrôle)⁸¹.

Certains auteurs ont pu relever⁸² qu'avoir la compétence et l'autorité pour intervenir sur le résultat proposé par l'algorithme ou le système d'IA ne signifie pas qu'il y a concrètement et dans tous les cas, intervention humaine.

En outre, selon le juriste Winston Maxwell, deux autres éléments sont indispensables pour que l'intervention humaine soit effective⁸³ :

- Avoir une connaissance du fonctionnement de l'algorithme et de ses limitations, c'est-à-dire savoir sur la base de quelles règles il a été conçu, à quel moment, etc. ;
- S'engager dans un processus de réflexion qui tient compte du contexte de la décision (faire preuve d'esprit critique vis-à-vis du fonctionnement du système et des résultats produits), se demander si on n'est pas dans un cas « atypique » au regard des données utilisées en entrée pour l'apprentissage de l'algorithme.

À défaut, l'intervention humaine pourrait n'être prévue qu'à titre purement formel. Prévoir des êtres humains dans la supervision risque alors « *de transformer les intéressés en 'amortisseurs moraux', en totems dont le rôle central deviendra de prendre la faute sur eux, même s'ils ne maîtrisent que partiellement le système, et susceptibles de servir de boucs émissaires aux entreprises et organisations cherchant à se dégager de leurs responsabilités* »⁸⁴. Il est donc nécessaire de déterminer les objectifs et les moyens assignés à l'intervention humaine.

De façon générale, deux types d'éléments, aujourd'hui largement documentés, conduisent à questionner la substance des interventions humaines dans nombre de décisions dites partiellement automatisées.

D'une part, dans les faits, le nombre massif des décisions contribue, dans certains cas, à s'interroger sur la réalité de l'intervention humaine dans des processus de prise de décisions administratives individuelles.

La substance de cette intervention est d'autant plus sujette à caution quand on sait que les algorithmes et autres systèmes d'IA sont souvent précisément mis en place pour accélérer et, du point de vue de l'organisme qui les déploie, simplifier la prise de décision et soulager les agents de tâches répétitives et chronophages, voire réaliser des économies de personnel.

LE CAS DE PARCOURSUP – LES ALGORITHMES LOCAUX UTILISÉS PAR DES ÉTABLISSEMENTS D'ENSEIGNEMENT SUPÉRIEUR POUR LE CLASSEMENT DES CANDIDATURES

Comme indiqué précédemment, la procédure Parcoursup est constituée de plusieurs étapes, dont l'une, essentielle, consiste dans le classement, par les établissements d'enseignement supérieur et plus précisément par les commissions d'examen des vœux (CEV)⁸⁵, des dossiers des candidatures reçues via la plateforme nationale.

Les établissements peuvent utiliser un algorithme, qui peut être soit l'outil d'aide à la décision (« OAD ») proposé par le ministère, soit un algorithme dit « maison ». Comme l'indique sa dénomination, l'outil algorithmique ne doit constituer qu'une aide à la décision car les décisions de classement prises sur le fondement de cet algorithme sont supposées être des décisions non entièrement automatisées⁸⁶. C'est d'ailleurs ce qu'a jugé le Conseil constitutionnel, saisi du dispositif, puisqu'il a établi d'une part, que l'usage des traitements algorithmiques par les établissements pour classer les dossiers n'est qu'une faculté et d'autre part que, lorsque les établissements ont recours à ces systèmes, « la décision prise sur chaque candidature ne peut être exclusivement fondée sur un algorithme » et qu'« elle nécessite, au contraire, une appréciation des mérites des candidatures par la commission d'examen des vœux, puis par le chef d'établissement »⁸⁷.

Le critère de l'intervention humaine est ici repris, via l'intervention de la CEV et du chef d'établissement – personnes qui ont effectivement la compétence pour modifier la décision. Mais la réalité et l'importance de ces interventions n'étant pas précisées, la qualification opérée par le Conseil constitutionnel peut, dans son application, être considérée par des auteurs comme « sujette à caution »⁸⁸.

Ce doute est en effet accru lorsqu'on considère la masse de dossiers de candidatures que les CEV doivent examiner. La doctrine⁸⁹ se montre dubitative sur la capacité des commissions à examiner dans un délai resserré (un mois environ, entre le moment où la CEV accède

aux dossiers et le moment où la décision doit être rendue) et mentionne qu'en 2018, et pour la seule licence de droit, l'université Paris I avait reçu 14 777 candidatures, Paris II 13 084 et Paris Descartes 9 841. Comme l'a relevé un sénateur devant la Commission de la culture, de l'éducation et de la communication du Sénat en 2019, à propos de l'université Paris Descartes, une part importante des dossiers de candidatures sont exclus lors du premier classement opéré par l'algorithme utilisé : pour tous ces dossiers, et les élèves concernés, on ne peut que constater l'absence d'intervention humaine⁹⁰. Récemment, une société savante d'informatique indiquait explicitement qu'il était recouru à « une procédure de classement automatisée des candidats »⁹¹.

Il ne s'agit pas ici de mettre en cause les membres des CEV, à qui revient une tâche en partie impossible qui rend parfois incontournables l'usage d'outils algorithmiques quantitatifs. Mais on ne peut que souligner le fort décalage entre le droit et les pratiques qu'il dessine, particulièrement critique ici, au regard de l'importance des décisions en matière d'admission dans l'enseignement supérieur pour les nombreuses personnes concernées, mineures pour une grande partie d'entre elles, chaque année.

LE CAS D'AFFELNET – L'ALGORITHME UTILISÉ POUR FACILITER LA GESTION DE L'AFFECTATION DES ÉLÈVES ET DES APPRENTIS DANS LES CLASSES DE SECONDE ET DE PREMIÈRE

La procédure d'affectation des élèves par le net (AFFELNET) permet chaque année à des centaines de milliers d'élèves en classe de troisième d'effectuer leurs vœux d'entrée au lycée, en seconde générale, technologique ou professionnelle et de bénéficier d'une affectation. Dans le cadre de cette procédure, chaque vœu est examiné en considération, notamment, des éléments suivants : le classement établi par la famille, les évaluations de l'élève (notamment ses résultats scolaires) et la capacité d'accueil des établissements. Chacun de ces éléments donne lieu à l'attribution d'un certain nombre de points, qui permettent d'établir le barème.



Dans ce cadre, un algorithme est utilisé pour « *faciliter la gestion de l'affectation des élèves* »⁹² dans chaque académie, en fonction des priorités et stratégies définies par le recteur. En d'autres termes, cet algorithme est censé servir à la prise de décision, mais non la fonder exclusivement. En outre, l'article D. 331-38 du code de l'éducation prévoit que « *L'affectation est de la compétence de l'inspecteur d'académie, pour les formations implantées dans le département. Il est assisté d'une commission dont la composition et le fonctionnement sont définis par arrêté du ministre chargé de l'éducation* » : comme pour le cas de Parcoursup, on retrouve des indications relatives à l'existence et à la compétence d'agents censés intervenir dans la prise de décision.

Or, dans le cadre de l'instruction d'une réclamation, le Défenseur des droits a pu constater que la décision d'affectation de l'élève concerné avait été prise de façon entièrement automatisée⁹³. Dans ce dossier, une décision d'affectation avait, en effet, été prise vis-à-vis d'une élève pour laquelle la fiche barème laissait apparaître la seule valeur de « 0 » dans la catégorie relative aux évaluations scolaires.

À l'issue du premier tour de la campagne d'affectation AFFELNET Lycée 2021, tous les vœux de l'élève avaient été refusés. C'est seulement parce que les parents, étonnés de ces refus au regard des résultats scolaires de leur fille, ont demandé à avoir accès aux informations sur la base desquelles la décision d'affectation avait été prise que la grossière erreur administrative a été identifiée et rectifiée. Dans ce dossier, le fait qu'une décision a été prise sur la base d'un barème qui fait apparaître la valeur de 0.000 dans la catégorie « évaluations », sans que le caractère exceptionnel de cette donnée amène la commission à procéder à des vérifications sur son caractère exact, constituait un élément fort conduisant à présumer que la décision avait été prise de manière entièrement automatisée. L'instruction du Défenseur des droits a permis de confirmer cette présomption : aucun élément sur l'intervention de la commission susmentionnée, ni sur les avis qu'elle aurait émis, n'a été communiqué par l'académie aux services de l'institution.

Si, selon l'académie, une telle donnée (0.000) pouvait correspondre à un cas de dispense ou d'absence de l'élève, le caractère exceptionnel d'une telle situation aurait dû, en toute hypothèse, l'amener à vérifier les raisons d'une telle donnée au moment où la décision était à l'étude. Or, tel n'a pas été le cas.

Le fait qu'un seul dossier a été porté à la connaissance du Défenseur des droits sur cette procédure ne doit pas conduire à considérer qu'il s'agit d'un cas isolé. Au contraire, les éléments mis en lumière par l'instruction laissent penser que les décisions peuvent, dans le cadre de cette procédure, avoir été et être encore massivement entièrement automatisées, en tout cas pour une partie des élèves.

Ces cas ne sont que des exemples, contribuant à interroger la substance de l'intervention humaine.

D'autre part, l'existence largement étayée de biais conduit à s'interroger sur la réalité de l'intervention humaine dans des processus de prise de décisions, ici administratives individuelles.

Considérer la possibilité d'une « intervention humaine significative » comme le critère de répartition entre d'une part des décisions individuelles entièrement automatisées et d'autre part les décisions prises sur le fondement d'un algorithme relève selon la doctrine⁹⁴ d'une vision technoneutre, largement critiquée. Dit autrement, la technique, créée par la volonté humaine, peut venir l'influencer. Comme le relève le penseur Jacques Ellul dans son ouvrage *Le système technicien* dès 1977, « en tant qu'hommes modernes nous ne sommes plus appelés à utiliser des techniques, mais à vivre avec des techniques et au milieu d'elles ». Il devient impossible d'imaginer un « homme souverain trônant dans cette collection et agissant sur la technique en toute indépendance ». Dans *Le bluff technologique* paru en 1988, Ellul indique que reconnaître que la technique n'est pas neutre, c'est reconnaître « qu'elle a son propre poids, ses propres déterminations, ses propres lois, autrement dit, en tant que système, elle impose sa logique ».

Dès 1956, le philosophe allemand Günther Anders considère, dans *L'obsolescence de l'homme*, qu'à partir du moment où l'on affirme que les techniques sont des moyens, on part d'un postulat inexact. Ces instruments ne sont pas des moyens car ils sont des décisions prises à l'avance, « décisions, précisément, qui sont prises avant même qu'on nous offre la possibilité de décider » : les « moyens eux-mêmes, les instruments eux-mêmes ne sont pas de simples objets que l'on peut utiliser, mais déterminent déjà, par leur structure et leur fonction, leur utilisation ainsi que le style de nos activités et de notre vie, bref, nous déterminent ».

Les analyses de ces auteurs révèlent à quel point le rôle de la volonté de l'homme dans l'utilisation de la technique est réduit : il n'est dès lors plus possible de considérer que tout objet technique est soumis à la volonté de la personne humaine, qui resterait quant à elle souveraine ; à l'inverse, il est essentiel de comprendre la façon dont le fonctionnement de l'objet technique soumet ou oriente cette volonté. La position dite techno-structuraliste appréhende ainsi la complexité des rapports entre la personne humaine et la technique, et permet de formaliser d'un point de vue théorique la critique du critère de répartition entre décisions automatisées et partiellement automatisées.

Au-delà des alertes sérieuses qui sont faites sur les **risques de perte d'esprit critique des agents publics face aux résultats des algorithmes et des systèmes d'IA avec lesquels ils doivent composer dans un contexte donné**⁹⁵, plusieurs types de biais doivent être soulignés⁹⁶ :

- **Le biais d'ancrage**, qui se manifeste par l'incapacité ou la difficulté de l'agent à se départir de la « première impression » que lui fournit le système, quand celui-ci intervient en amont ;
- **Le biais d'automatisation**, qui se définit comme la tendance de l'humain à faire davantage confiance aux résultats produits par la machine qu'à son propre jugement et, ainsi, à avaliser de façon systématique ou excessive les recommandations qu'elle formule.

Comme le relève le Conseil d'État, le taux d'approbation humaine des résultats de la machine constitue « *l'un des principaux indices du biais d'automatisation* »⁹⁷ (avec un taux d'approbation à 100%, on peut sérieusement douter de l'intervention de l'agent). Mais une évaluation statistique s'avère *a priori* insatisfaisante pour jauger la substance de l'intervention humaine et seule une approche globale du contexte et des conditions de la prise de décision apparaît pertinente.

À cet égard, le Conseil d'État proposait dès 2014⁹⁸ de « *confirmer que l'intervention humaine dans la décision doit être réelle et pas seulement formelle* » et suggérait d'« *indiquer dans un instrument de droit souple les critères d'appréciation du caractère effectif de l'intervention humaine* ». Cette recommandation peine encore à se concrétiser.

- De possibles **biais induits par le régime de responsabilité**⁹⁹, lequel peut conduire la personne en charge du contrôle humain à « *suivre excessivement les recommandations algorithmiques* ». La raison en est que se tromper du fait d'avoir suivi le résultat ou la recommandation algorithmique serait considéré comme moins fautif que se tromper en allant contre le résultat ou la recommandation du système. Selon Maxwell, « *ce biais se retrouvera dans toute situation où l'humain devra justifier, à son supérieur hiérarchique par exemple, le choix de ne pas suivre une recommandation algorithmique alors que le choix inverse ne nécessiterait aucune justification de la part du décideur humain* ». Or, le régime de la responsabilité de l'agent administratif en charge de la décision n'apparaît pas clair dans ce cadre.

L'ensemble de ces constats conduit à s'interroger sur la portée réelle **de la distinction** entre décisions partiellement automatisées d'une part et décisions entièrement automatisées d'autre part. Cette distinction prend toute son importance quand on sait que le cadre juridique :

- **Repose sur un droit, celui de ne pas faire l'objet d'une décision fondée exclusivement** sur un traitement automatisé produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire¹⁰⁰. Si, comme on l'a vu, il est assorti **d'exceptions, dont une applicable dans le champ de l'action administrative**, il constitue néanmoins **un principe**.
- A prévu, pour les décisions administratives individuelles entièrement automatisées, des garanties : principalement l'interdiction d'utiliser des données personnelles dites sensibles pour fonder des décisions individuelles entièrement automatisées.

En outre, la façon dont le traitement algorithmique contribue à la prise de décision fait partie des informations que l'administration doit communiquer à la personne concernée, sur demande de cette dernière, lorsqu'un algorithme a servi à prendre une décision administrative individuelle, que celle-ci soit entièrement ou partiellement automatisée. Un lien direct existe donc entre la clarification de la substance de l'intervention humaine et l'information donnée à cet égard aux usagers qui en font la demande (cf. partie III).

AU REGARD DE CES CONSIDÉRATIONS, LA DÉFENSEUR DES DROITS RECOMMANDE :

- **Dans les cas où la décision administrative individuelle prise sur le fondement du résultat d'un algorithme ou d'un système d'IA est qualifiée de décision partiellement automatisée : d'édicter des critères et des modes opératoires obligatoires, alternativement ou cumulativement, pour qualifier plus précisément « l'intervention humaine ».**

Des travaux conduits par des élèves de l'ENA¹⁰¹ permettent par exemple d'identifier plusieurs éléments :

- Le moment de l'intervention humaine ;
- La forme que revêt cette intervention humaine :
 - > Vise-t-elle à valider ou infirmer le résultat produit ? Dans ce cas, s'agit-il d'une décision discrétionnaire ou d'une vérification de l'exactitude de ce résultat ?¹⁰²

- > Quels autres éléments que le résultat produit par le système l'intervention humaine prend-elle en considération ?

En ce sens, plusieurs propositions¹⁰³ suggèrent de définir à l'avance « *des tâches spécifiques pour l'humain et pour l'ordinateur* », ce qui irait dans le sens des recommandations du Conseil de l'Europe sur la responsabilité¹⁰⁴. La « *définition précise de la mission du contrôleur humain* » et l'identification des moyens alloués pour ce faire (en termes matériels, humains, temporels, etc.) permettraient de recentrer la responsabilité et de formaliser concrètement l'intervention humaine. Et la programmation de « *tests d'efficacité* »¹⁰⁵ de cette intervention humaine assurerait de son effectivité.

Ces éléments devraient être définis en amont de la mise en œuvre de l'algorithme ou du système d'IA par l'administration concernée. Ils constitueraient également un moyen de répondre à l'obligation posée par le CRPA en cas de demande d'information complémentaire de la personne concernée. En effet, dans ce cas, le service public doit notamment indiquer à la personne concernée « *le degré et le mode de contribution du traitement algorithmique à la prise de décision* »¹⁰⁶ (cf. partie III).

• Faute de critères qualifiant l'intervention humaine, envisager une autre méthode de répartition.

Il s'agirait d'évaluer l'opportunité de se fonder sur le critère de « *l'applicabilité directe* » du résultat de l'algorithme ou du système d'IA.

Le résultat algorithmique est dit d'applicabilité directe lorsqu'il se suffit à lui-même au sens où il est à lui seul suffisant pour produire la substance d'une décision administrative individuelle. Un faisceau d'indices est mobilisé pour caractériser en ce sens ce résultat qui doit être :

- clair : il ne doit pas être nécessaire de l'interpréter ;
- précis : il doit être détaillé, applicable à la situation évaluée ;
- complet : il ne doit pas laisser de marge d'appréciation.

Selon Liane Huttner, à l'origine de cette proposition¹⁰⁷, ce n'est pas l'algorithme qui est classé dans une catégorie mais son usage à un moment donné pour une décision déterminée : par conséquent, un même algorithme peut être qualifié différemment en fonction de l'utilisation qui en est faite dans le processus décisionnel.

Sur cette base, il est proposé de distinguer :

- D'une part, les algorithmes de prise de décision qui permettent de fonder entièrement une décision ;
- D'autre part, les algorithmes d'aide à la décision qui ne permettent pas de fonder entièrement une décision mais qui peuvent l'éclairer.

> Le critère de répartition consiste dans « l'applicabilité directe » du résultat de l'algorithme ou du système d'IA.

En cas d'utilisation d'un algorithme ou d'un système d'IA dont le résultat est « *d'applicabilité directe* », il conviendrait ainsi d'appliquer automatiquement le régime prévu pour les décisions administratives individuelles entièrement automatisées.

Il convient ici d'avoir à l'esprit que le cadre juridique interdit la prise de décisions administratives individuelles entièrement automatisées sur la base d'un traitement de données à caractère personnel contenant des données sensibles au sens du RGPD. Or, pour ne prendre que les données concernant la santé, et en particulier celles relatives à la situation de handicap, elles sont présentes dans un grand nombre de traitements du secteur social notamment. Une réflexion doit donc se tenir sur le périmètre des données sensibles et ce à l'échelle européenne compte tenu du RGPD et du rôle de la jurisprudence de la CJUE en la matière.

La Défenseure des droits recommande, dans ce contexte, que des lignes claires soient fixées par le Comité européen sur la protection des données (CEPD) sur la définition des données sensibles au sens du RGPD et de la loi « *Informatique et Libertés* » et que soit engagée une réflexion sur la pertinence de cette interdiction, en lien avec les risques de discrimination. Le CEPD est un organe de l'UE, indépendant, dont le rôle est de veiller à ce que le RGPD soit appliqué de manière cohérente dans les pays de l'UE.

PARTIE III

LA TRANSPARENCE : UN PRINCIPE FORT, MAIS EN PRATIQUE ?

La transparence de l'action administrative n'est pas une question nouvelle, et précède largement celle de l'algorithmisation de l'action publique. La tension entre le pouvoir discrétionnaire de l'administration d'une part et les droits associés qui se sont développés au fil au temps en faveur de la transparence d'autre part en témoigne. Comme l'indiquait Jean-Marc Sauvé¹⁰⁸ : « *La transparence permet la mise en œuvre effective des principes et des valeurs qui fondent l'action publique et elle renforce ce faisant la confiance des citoyens dans les institutions : elle en est même un pilier. (...) Elle participe à la mise en œuvre effective du principe d'égalité. Égalité d'accès aux emplois publics par exemple, pour prévenir toute forme de népotisme ou de favoritisme, ou égalité des usagers du service public, pour faire échec aux privilèges liés à l'asymétrie des informations. Au travers, notamment, du droit d'accès aux documents administratifs et de l'obligation de motivation, la transparence favorise également l'impartialité et l'objectivité du service public. C'est pourquoi elle est, dans ses différentes déclinaisons juridiques, une composante du droit à une bonne administration* »¹⁰⁹.

Parmi les décisions actuellement prises par ou avec l'aide d'algorithmes, il existe des décisions administratives qui ne faisaient pas toujours l'objet d'une motivation claire et pouvaient être prises selon des procédures opaques. Ici, c'est justement l'utilisation de systèmes algorithmiques ou d'IA qui peut apporter de la transparence, parce que leurs règles de fonctionnement doivent être -un minimum- définies.

En vertu de l'article 47 de la loi « Informatique et Libertés », l'information des usagers ayant fait l'objet de décisions administratives automatisées constitue une condition de validité de ces décisions¹¹⁰. Encore faut-il être en mesure d'identifier précisément ce que doivent être la nature et le niveau de cette information pour que cette obligation soit effective¹¹¹. Se pose également la question des informations sur les traitements algorithmiques opérés dans les services publics, qui doivent être portées plus largement à la connaissance de la population et des usagers. Puis, au-delà des limites légales qui encadrent cette information et des difficultés pratiques rencontrées, comment faire en sorte que ce droit à l'information se concrétise pleinement pour les usagers du service public, et pour quoi faire ?



1· DES ENJEUX MAJEURS

La transparence sur les traitements automatisés utilisés par l'administration et sur les décisions prises sur ce fondement a été prônée pour équilibrer le cadre juridique, de façon cohérente avec l'idée qui fait de la transparence une « *forme de rationalisation, de surveillance et surtout de légitimation de l'action publique* »¹¹². Le droit à l'information dont disposent les usagers du service public dès qu'un algorithme participe à la prise d'une décision administrative individuelle qui les concerne découle ainsi du principe reconnu à l'article 15 de la Déclaration des Droits de l'Homme et du Citoyen selon lequel « *La société a le droit de demander compte à tout agent public de son administration.* ». Il procède également de l'obligation de motivation de certaines décisions administratives prévues à l'article L. 211-5 du CRPA¹¹³.

Si la communication, à l'utilisateur concerné, d'informations relatives à l'algorithme qui a servi à prendre la décision administrative individuelle paraît primordiale pour assurer l'effectivité d'un principe de rang constitutionnel, il faut considérer que la transparence est aussi ce qui permet de comprendre certaines décisions et par conséquent de pouvoir ensuite en débattre et les contester utilement. La transparence constitue ainsi un élément clef de la responsabilité de l'action administrative. En ce sens, elle doit ainsi être envisagée comme le prérequis pour lutter contre d'éventuels erreurs, abus et discriminations. Certains auteurs parlent à ce sujet de « *loyauté* », concept qui englobe la transparence dont elle n'est qu'une facette : « *la loyauté suppose que l'algorithme dise ce qu'il fait mais aussi qu'il fasse ce qu'il dit et que ce qu'il dit corresponde à la satisfaction de l'intérêt général, dans le respect du droit (...). En somme, la loyauté implique que l'algorithme soit transparent, intègre et légal* »¹¹⁴.

L'algorithmisation de l'administration contribue donc à renouveler l'impérative transparence de l'action publique.

Plusieurs phases peuvent être distinguées, chacune d'entre elles participant à faire du droit à l'information une réalité¹¹⁵ :

- L'information vise à « rendre visible », montrer, porter à la connaissance ;
- L'explication vise à « rendre compréhensible » - étant précisé qu'on peut distinguer les explications ponctuelles (pour expliquer une décision ou un résultat particulier) des explications globales (relatives à la logique générale du système) ;
- La justification vise à « rendre acceptable » - elle fait ainsi appel à des éléments extérieurs tels que les textes de loi et les objectifs de l'action publique pour justifier le résultat¹¹⁶.

Si, comme nous le verrons, **certaines mentions relatives à la transparence sont prévues à la seule intention de l'utilisateur concerné, d'autres obligations ont une vocation plus collective** : ainsi en est-il par exemple de la publication en ligne à laquelle sont tenues certaines administrations s'agissant des règles définissant les principaux traitements algorithmiques utilisés dans l'accomplissement de leurs missions lorsque ces traitements fondent des décisions individuelles. À l'enjeu du caractère compréhensible des informations communiquées au seul usager, répond l'enjeu du caractère accessible des informations publiques afin que divers acteurs¹¹⁷ puissent s'en emparer, participer à la compréhension de l'impact de ces systèmes sur la société¹¹⁸ et « *comprendre, contester et discuter le rôle de la machine dans la décision publique* »¹¹⁹. Cette appropriation collective participerait à la pleine compréhension pour l'utilisateur qui ne détient quasiment jamais les compétences pour expertiser l'algorithme. Ceci alors qu'une telle expertise se révèle indispensable pour identifier d'éventuels effets contestables voire discriminatoires, effets qui ne seraient aperçus qu'à l'échelle collective, à partir d'une analyse du code source et/ou des résultats agrégés produits par le système¹²⁰.

Or, ce qui est un droit pour les usagers et pour les acteurs engagés sur ces sujets constitue aussi, pour l'administration elle-même **un enjeu d'intelligibilité de sa propre action**¹²¹, voire de (re)prise en main des outils et donc de responsabilisation.

Il apparaît en effet essentiel de « *remettre les administrations au centre de l'expertise* » et d'éviter qu'elles « *utilisent des outils sans en comprendre le fonctionnement, ce qui les mettrait en position de dépendance vis-à-vis des résultats issus de l'application d'algorithmes* »¹²².

Il s'agit d'une question fondamentale dans un contexte où l'action publique recourt de plus en plus à de tels systèmes informatiques, en s'appuyant fréquemment sur des prestataires privés, et où « *le rythme auquel les algorithmes doivent imprimer les changements législatifs et réglementaires est un facteur déstabilisant pour les agents des directions des systèmes d'information* »¹²³. Si la compétence du Défenseur des droits concerne les droits des usagers des services publics et non le fonctionnement général de ces services¹²⁴, l'un et l'autre paraissent ici liés. Il s'agit ainsi de faire en sorte que, par son fonctionnement interne, et par le développement de la capacité des agents à comprendre les systèmes utilisés, les droits des usagers soient respectés. Comme le souligne Liane Huttner, l'auteur de la décision est, tout autant que son destinataire, concerné par les différents mécanismes visant à soumettre l'algorithme et la décision prise à un processus contrôlé et contestable : l'agent doit pouvoir comprendre le fonctionnement du système et intervenir pour en contrôler les résultats¹²⁵.

Par exemple, la transparence d'un code source de bonne qualité, c'est-à-dire sans obfuscation¹²⁶, doit aussi constituer pour l'administration et le service informatique concerné au premier chef, un « *objectif désirable (...) puisque ces techniques participent de la bonne ingénierie logicielle et de la qualité de maintenance des systèmes d'information* »¹²⁷ : un objectif à mettre en regard avec l'obligation (dans certaines conditions) de publication du code source utilisé par l'administration.

FOCUS

L'ENJEU DE L'EXPLICABILITÉ DES SYSTÈMES D'IA OU « XAI »

Les systèmes d'IA et les algorithmes en général sont parfois opaques et leurs résultats peuvent susciter des difficultés d'interprétation. Or, ainsi que l'indique le règlement de sur l'IA de l'UE¹²⁸ : *« l'exercice de droits fondamentaux procéduraux importants, tels que le droit à un recours effectif et à accéder à un tribunal impartial, ainsi que les droits de la défense et la présomption d'innocence, pourrait être entravé, en particulier lorsque ces systèmes d'IA ne sont pas suffisamment transparents, explicables et documentés ».*

Selon la définition de la Cnil¹³⁰ : *« Dans le domaine de l'intelligence artificielle, l'explicabilité est la capacité de mettre en relation et de rendre compréhensible les éléments pris en compte par le système d'IA pour la production d'un résultat. Il peut s'agir, par exemple, des variables d'entrée et de leurs conséquences sur la prévision d'un score, et ainsi sur la décision. Les explications doivent être adaptées au niveau de compréhension de la personne auxquelles elles sont destinées. ».*

Cette question n'est pas nouvelle dans le domaine de l'IA. Néanmoins, comme le soulignent certains¹³⁰, elle est devenue centrale avec le développement et l'importance de systèmes appelés « boîtes noires ». A noter que, selon les mêmes auteurs, l'usage de cette expression renvoie à deux cas distincts : d'une part, celui d'un système dit « propriétaire » pour lequel il n'existe pas notamment d'accès au code source pour les personnes qui ne disposent pas de la licence ; d'autre part, celui d'un système à la complexité telle que sa compréhension n'est pas acquise.

Une communauté de recherche s'est développée au niveau mondial ces dernières années sur le thème de l'explicabilité des systèmes, appelé xAI (pour *explainable AI* c'est-à-dire l'explicabilité de l'IA)¹³¹.

Parmi de nombreux chercheurs, Daniel Le Métayer et Clément Hénin¹³² s'intéressent ainsi aux différents moyens de s'attaquer à l'opacité des systèmes algorithmiques.

De nombreuses méthodes sont étudiées, sans faire consensus à ce jour comme le relève la Cnil¹³³. Certaines fonctionnent en « boîte noire », c'est à dire sans connaissance du code du système, alors que d'autres, dites en « boîte blanche », interviennent sur ce code. Certaines ont pour objectif de rendre les systèmes interprétables (ou parle alors de système « intelligible ») alors que d'autres cherchent à produire des explications *a posteriori* de systèmes opaques.

Si chaque approche comporte des avantages et des inconvénients, plusieurs défis doivent encore être relevés.

Fournir des explications ou des justifications véritablement compréhensibles par des utilisateurs profanes (médecin, juriste ou personne affectée par les résultats des systèmes, par exemple)¹³⁴.

En général, les explications fournies par les méthodes existantes sont unilatérales, c'est-à-dire données par le responsable à la personne concernée, sans échange, et fournies sous une forme fixe, préalablement décidée (par exemple des arbres de décision ou des listes de facteurs prépondérants) alors que les utilisateurs ont des besoins variés selon leurs motivations et de leur degré de compétence.

Les justifications et les contestations n'ont pas fait l'objet de nombreux travaux de recherche jusqu'à ce jour. L'évaluation expérimentale des résultats représente une question qui mérite plus d'attention : il est plus simple de formaliser des indicateurs de performance du système et d'évaluer la performance de ce système, que de l'expliquer et de mesurer la qualité de l'explication¹³⁵.

2· DES DROITS D'ORES ET DÉJÀ EN VIGUEUR

Deux cadres juridiques distincts portent des exigences en matière de transparence des décisions administratives individuelles entièrement ou partiellement automatisées.

D'un côté, le cadre juridique relatif à la protection des données personnelles, dont le respect est contrôlé en premier lieu par la Cnil¹³⁶ ;

De l'autre, le cadre juridique relatif aux relations entre le public et l'administration. La Cada est l'autorité référente s'agissant du respect de la liberté d'accès aux documents administratifs et aux archives publiques ainsi qu'à la réutilisation des informations publiques¹³⁷, mais n'est pas compétente pour contrôler le respect de certains droits relevant de la transparence et mentionnés dans le CRPA.

La transparence de l'action administrative déborde le sujet des algorithmes et des systèmes d'IA utilisés vis-à-vis des usagers : le droit d'accès consacré par le CRPA concerne tous types de documents (cf. article L. 300-2 du CRPA).

À ce sujet, des propositions de réforme émergent dans le débat public autour du droit de communication des documents administratifs en général : allongement du délai (1 mois aujourd'hui) dont disposent les administrations pour répondre aux demandes de communication avant que la Cada puisse être saisie pour avis, transformation des avis de la Cada en avis conformes qui s'imposeraient aux administrations (ce qui n'est pas le cas aujourd'hui), augmentation de son budget. Le présent rapport, qui se concentre sur l'action publique algorithmique, n'ambitionne donc pas de traiter ces enjeux de façon globale.

Concernant le cadre juridique applicable, il faut comprendre :

- Que le cadre « données personnelles » s'applique aux seules décisions (en réalité très nombreuses) prises (par l'administration ou par une entreprise privée) sur le fondement d'un traitement de données à caractère personnel et, dans le cadre de l'action administrative algorithmique, que ses dispositions relatives à l'explication sur la logique de l'algorithme ne s'appliquent que lorsque la décision est entièrement automatisée ou qu'est établi un profilage au sens de l'article 22 du RGPD ¹³⁸ ;
- Que le cadre « CRPA » s'applique de façon plus large : dès qu'une décision administrative est en jeu (qu'elle soit entièrement ou partiellement automatisée, qu'elle soit prise sur la base d'un traitement de données à caractère personnel ou non) et donc, dans de nombreux cas, de façon cumulative avec le cadre « données personnelles » ;
- Que le contenu des obligations respectives n'est pas identique mais se rejoint globalement.

Les tableaux récapitulatifs suivant proposent une vue des obligations issues de ces deux cadres juridiques, en tant qu'elles concernent la transparence sur les décisions administratives individuelles partiellement ou entièrement automatisées.

AU TITRE DU CRPA

➤ Que la décision administrative individuelle soit partiellement ou entièrement automatisée, qu'elle ait été prise ou non sur le fondement d'un traitement de données à caractère personnel, qu'elle concerne une personne physique ou une personne morale.

• Informations à fournir à la personne concernée par la décision administrative individuelle :

Il s'agit d'un mécanisme à deux étages :

*1^{er} étage*¹³⁹: la décision individuelle doit comporter une **mention explicite informant l'intéressé que ladite décision a été prise sur le fondement d'un algorithme** et qu'il ou elle peut obtenir des informations supplémentaires en en faisant la demande

! Exception si cela porte atteinte à des secrets protégés¹⁴⁰.

*2^e étage*¹⁴¹: si la personne concernée demande d'autres informations, l'administration doit lui communiquer sous une forme intelligible 4 séries d'informations relatives au traitement algorithmique sur la base duquel la décision a été prise (et dans la limite des secrets protégés par la loi) :

1. Le degré et le mode de contribution du traitement algorithmique à la prise de décision
2. Les données traitées et leurs sources ;
3. Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé ;
4. Les opérations effectuées par le traitement.

• Information générale à effectuer via une publication en ligne sur le site de l'administration concernée :

Conformément à l'article L. 312-1-3 du CRPA , les administrations¹⁴² de plus de 50 ETP¹⁴³ doivent publier en ligne les règles définissant les principaux traitements algorithmiques utilisés dans l'accomplissement de leurs missions lorsqu'ils fondent des décisions individuelles.

Une exception existe si cela porte atteinte à des secrets protégés¹⁴⁴.

• Accès et publication du code source via une demande dite « Cada » :

Le CRPA précise que le code source¹⁴⁵ constitue un document administratif¹⁴⁶.

À ce titre, il fait, comme les comptes rendus, procès-verbaux, rapports, études, etc., partie des informations que toute personne peut, dans le cadre posé par ce code, demander à l'État, aux collectivités territoriales, aux autres personnes de droit public ou aux personnes de droit privé chargées d'une mission de service public, dans le cadre d'une demande dite « Cada »¹⁴⁷.

Le code source est également soumis à une obligation de publication en ligne si la demande en est faite à l'administration¹⁴⁸ et à terme par défaut¹⁴⁹ (cette dernière obligation ne s'applique que pour les collectivités territoriales de plus de 3500 habitants et pour les administrations de plus de 50 ETP¹⁵⁰).

• De façon générale, accès à un document administratif concernant l'algorithme utilisé, via une demande dite « Cada » :

Il est possible de s'appuyer classiquement sur le droit de communication pour obtenir tout document administratif préparatoire ou faisant le bilan des résultats de l'algorithme (sauf si secret protégé).

AU TITRE DU RGPD ET DE LA LOI « INFORMATIQUE ET LIBERTÉS »

- Quand la décision administrative est entièrement automatisée et qu'elle a été prise sur le fondement d'un traitement de données à caractère personnel et qu'elle concerne une personne physique

Modalités de la transparence :

L'article 12 du RGPD détaille les modalités des informations à fournir à la personne concernée (information à fournir de façon concise, transparente, compréhensible et aisément accessible).

Informations à communiquer à la personne concernée :

1^{er} étage : selon le cas, l'article 13 ou l'article 14 du RGPD détaille le contenu des informations à fournir d'emblée (et non sur demande de sa part) à la personne concernée sur le traitement qui est fait de ses données personnelles.

En cas de décision entièrement automatisée, y compris un profilage, l'organisme responsable doit communiquer à la personne concernée des informations utiles concernant la logique sous-jacente et l'importance et les conséquences prévues de ce traitement pour la personne concernée (articles 13.2-f et 14.2-g du RGPD).

2^e étage : en cas de demande de la personne concernée, le responsable du traitement 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* » (article 47 de la LIL).

3^e étage : la personne concernée peut en outre exercer son droit d'accès¹⁵¹ (la confirmation que ses données personnelles sont traitées et un certain nombre d'informations à ce sujet).

Demande d'accès au registre des traitements :

Chaque responsable de traitement doit, sauf exceptions, tenir un registre des activités de traitement avec un certain nombre d'informations obligatoires¹⁵² (finalités du traitement, description des catégories de données personnelles concernées, catégories de destinataires des données personnelles, etc.).

Ce registre de traitement, qui peut permettre de comprendre les algorithmes utilisés, peut être demandé par une demande dite « Cada ».

Il ressort donc clairement - des dispositions du CRPA en particulier - que les obligations posées sont envisagées à la fois d'**un point de vue individuel** (pour l'utilisateur concerné) et d'**un point de vue collectif** (pour les tiers que sont les associations, les équipes de recherche, les personnes accompagnant des usagers, tout citoyen intéressé ou collectif de citoyens, etc.).

FOCUS

LES EXIGENCES DE TRANSPARENCE ET LEURS CONSÉQUENCES SUR LES ALGORITHMES QUI PEUVENT ÊTRE UTILISÉS POUR LES DÉCISIONS ADMINISTRATIVES INDIVIDUELLES

- Le Conseil constitutionnel, conduit à se prononcer, en 2018, sur la prise de **décisions administratives individuelles entièrement automatisées**, a considéré que « *le responsable du traitement doit s'assurer 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. Il en résulte que ne peuvent être utilisés, comme fondement exclusif d'une décision administrative individuelle, des algorithmes susceptibles de réviser eux-mêmes les règles qu'ils appliquent, sans le contrôle et la validation du responsable du traitement* »¹⁵³. Le Conseil constitutionnel est donc parti de l'impératif de transparence pour **poser une limite forte aux systèmes susceptibles de fonder une décision entièrement automatisée**.
- Cette limite semble pouvoir être **étendue aux systèmes entrant en œuvre dans la prise de décisions partiellement automatisées**. En effet, l'information que le service public doit fournir à la personne concernée sur demande

de cette dernière lorsqu'un système a été utilisé pour fonder (partiellement ou entièrement) une décision comprend, comme on l'a vu, « *les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé* »¹⁵⁴.

- Selon la Cnil, « *le paramètre est la propriété apprise des données utilisées pour l'entraînement (par exemple le poids de chaque neurone d'un réseau)* »¹⁵⁵.
 - Or, il n'est pas évident que, pour un modèle issu de l'apprentissage automatique, la pondération exacte des paramètres soit déterminable précisément. Selon le juriste Winston Maxwell, celle-ci « *restera un mystère* »¹⁵⁶. S'il considère qu'il sera possible, grâce à certaines méthodes d'explicabilité, de « *visualiser les poids approximatifs des différents paramètres* », il insiste sur le caractère approximatif de cette estimation.
- > La conséquence directe de ce constat est que l'utilisation d'algorithmes apprenants pour fonder, même partiellement, des décisions administratives individuelles doit être interrogée au regard des dispositions réglementaires en vigueur dont l'exigence de transparence.

S'agissant de l'évolution des textes, le règlement sur l'IA de l'UE prévoit explicitement un « droit à l'explication des décisions individuelles » pour les personnes concernées par une décision prise par un organisme qui déploie un système d'IA considéré à haut risque et qui produit des effets (juridiques ou significatifs) sur elles¹⁵⁷. Dans ce cas, la personne a le droit d'obtenir « *des explications claires et pertinentes sur le rôle du système d'IA dans la procédure décisionnelle et sur les principaux éléments de la décision prise* ».

Au-delà du fait que le règlement sur l'IA ne fait, à ce niveau, **pas de distinguo entre décision entièrement automatisée** et décision partiellement automatisée, il est significatif de relever que ces dispositions du texte européen font état d'un critère de « pertinence », sans que soient précisés à ce stade les ressorts de celle-ci.

3· DES LIMITES JURIDIQUES, UNE RECOMMANDATION POUR DES AMÉLIORATIONS

D'un point de vue juridique, les limites à la transparence et à la communication des informations ont trait :

1. Au fait que **le cadre juridique applicable est parfois double** : en effet, en présence d'une décision administrative individuelle entièrement automatisée prise sur le fondement d'un traitement de données à caractère personnel, l'organisme auteur de cette décision doit respecter les obligations d'information du CRPA d'une part et de la loi « Informatique et Libertés » d'autre part.

EN CONSÉQUENCE, LA DÉFENSEURE DES DROITS RECOMMANDE :

- aux pouvoirs publics, de conduire une réflexion pour simplifier les obligations à la charge des organismes au titre de la transparence qui permettrait d'améliorer l'information à destination de l'utilisateur concerné, sans conduire à un amoindrissement de ses droits.

2. À la protection de certains secrets :

- Il peut s'agir d'informations non communicables dans les domaines directement listés dans le CRPA¹⁵⁸, notamment « la sécurité publique, la sécurité des personnes » ou « la sécurité des systèmes d'information des administrations » ou encore « la recherche et à la prévention, par les services compétents, d'infractions de toute nature »¹⁵⁹ ; l'appréciation des contours de ces secrets et de leur application au cas d'espèce relève de la compétence de la CADA :

> Ainsi, concernant le secret lié à la « sécurité des systèmes d'information des administrations », la CADA indique qu'elle apprécie « très strictement cette réserve, qui ne se présume pas et doit être démontrée à partir d'éléments circonstanciés fournis par les autorités administratives »¹⁶⁰ ;

> La CADA a refusé de faire droit à une demande de communication de l'algorithme du modèle de *data mining*

utilisé par la Cnaf pour attribuer un score de risque aux allocataires, afin de ne pas porter atteinte à la politique de lutte contre la fraude sociale menée par l'organisme et à l'efficacité des contrôles. Elle a en effet considéré que dévoiler les variables prises en compte et les coefficients associés reviendrait à divulguer les critères utilisés pour cibler les contrôles et ainsi à augmenter le risque de fraude individuelle ou organisée. La CADA a en revanche fait droit à la demande de communication des variables utilisées dans les modèles qui n'étaient plus en vigueur et sur le fondement desquels ne reposaient pas les contrôles actuels ou futurs¹⁶¹.

- Il peut s'agir d'informations listées dans des lois spéciales, comme c'est par exemple le cas pour le dispositif Parcoursup : pour « la nécessaire protection du secret des délibérations des équipes pédagogiques chargées de l'examen des candidatures », l'article L. 612-3 du code de l'éducation exclut l'application de deux articles du CRPA relatifs à la communication et à la publicité des traitements algorithmiques utilisés comme fondement, exclusif ou partiel, d'une décision administrative individuelle, ainsi que l'a relevé et validé le Conseil constitutionnel dans sa décision sur le sujet¹⁶².

3. Au nécessaire respect d'autres droits et en particulier :

- **Du droit au respect de la vie privée** : ainsi, un certain nombre de documents ne sont communicables qu'à l'intéressé¹⁶³.

> La CADA a refusé de faire droit à la demande de communication des modèles d'apprentissage entraînés de l'outil de pseudonymisation des décisions de justice développé par la Cour de cassation. Elle a estimé qu'elle serait en effet de nature à remettre en cause la finalité même de l'outil, à savoir sécuriser la diffusion publique des décisions de justice en assurant le respect de la vie privée des personnes concernées, du fait du risque d'opérations qui permettraient de reconstituer les données occultées. En ce sens, la CADA a estimé que la divulgation de ces documents mettrait en cause la protection de la vie privée¹⁶⁴ ;

- **Du droit à la propriété intellectuelle** : si le code source d'un algorithme constitue un document administratif communicable, en pratique, dans le cadre d'un système ayant fait l'objet d'un marché public, le régime de la propriété intellectuelle peut faire obstacle à sa communication. Le Conseil d'État considère en effet qu'un document dont l'administration n'est pas propriétaire ne peut être communiqué qu'avec l'accord de celui qui détient le droit de propriété¹⁶⁵.

4. Au caractère non contraignant de certaines procédures :

- La loi ne prévoit pas de **sanctions en cas de non-respect par les administrations et autres organismes concernés¹⁶⁶ des obligations** de transparence précitées du CRPA, sauf pour ce qui concerne la nullité de la décision en cas d'absence de la mention individuelle d'office. En particulier, l'obligation, sauf exception liée aux secrets protégés, de publier les « règles définissant les principaux traitements algorithmiques » au sens de l'article L. 312-1-3 du CRPA, n'est assortie d'aucune sanction.
- **Le caractère non liant des avis rendus par la Cada** : en effet, quand elle est saisie après rejet ou silence de l'administration à laquelle est demandé l'accès à un document administratif, la Cada peut rendre un avis favorable à la communication mais ne peut l'imposer à l'administration concernée. Il est ainsi parfois nécessaire, pour obtenir l'accès à un document administratif, de saisir le tribunal administratif compétent.

4- DES LIMITES PRATIQUES ET DES RECOMMANDATIONS POUR Y REMÉDIER

D'un point de vue pratique, la mise en œuvre du droit à l'information prévu par le CRPA et du principe de transparence rencontre des limites et bute sur des difficultés. Pour chacun, la Défenseure des droits formule une ou plusieurs recommandations afin de permettre de concrétiser le droit à l'information des usagers des services publics.

PLUSIEURS TYPES DE LIMITES SONT IDENTIFIÉS :

1. Un défaut de publication des « règles définissant les principaux traitements algorithmiques ».

Aujourd'hui et depuis l'entrée en application de **cette obligation prévue à l'art. L. 312-1-3 du CRPA**, plusieurs acteurs¹⁶⁷ font le constat, que son « effectivité est douteuse », certains allant même jusqu'à dire qu'elle est globalement restée « lettre morte » depuis 2016¹⁶⁸, à quelques exceptions près¹⁶⁹.

Le défaut de sanction du non-respect de cette obligation peut contribuer au manque d'entrain des administrations concernées. Cette publication, par exemple par différentes collectivités locales de même type (régions, par exemple), pourrait en outre permettre de comparer et discuter les pratiques et les choix politiques retenus – lesquels ne sont pas toujours explicités par ailleurs.

La Défenseure des droits salue la mise en place d'observatoires citoyens locaux et nationaux (parmi lesquels l'association Ouvre-Boite¹⁷⁰, le projet d'observatoire des algorithmes publics – ODAP¹⁷¹), qui ne doit cependant pas occulter la responsabilité des administrations concernées à cet égard.

EN CONSÉQUENCE, LA DÉFENSEURE DES DROITS RECOMMANDE :

- **Au Gouvernement, d'introduire une sanction en cas de non-respect de cette obligation et de mettre en place un recensement à l'échelle nationale relatif au respect de ces obligations de publication ;**
- **Aux administrations et organismes concernés, de procéder sans attendre aux publications obligatoires prévues par le CRPA ;**



• *En parallèle, la mise en place et le soutien d'observatoires citoyens locaux ou de recensements à l'échelle nationale).*

2. Une **difficile gouvernance**, au sein des services publics, de l'application du cadre juridique, relevant d'une diversité d'acteurs, souvent sans qu'aucun ne soit forcément désigné responsable (collaborateur délégué à la protection des données (DPO) – personne responsable de l'accès aux documents administratifs (PRADA), parfois peu connue¹⁷² – administrateurs ministériels des données, des algorithmes et des codes sources (AMDAC) et des **difficultés pratiques pour les administrations, collectivités et autres organismes concernés pour respecter les obligations existantes**, qui impliquent notamment de :

- **Définir une mention individuelle** à apposer sur la décision administrative individuelle concernée au sens du CRPA, qui plus est claire et accessible aux usagers ;
- **Savoir quelles informations supplémentaires** communiquer à la personne concernée au titre d'une réponse à une demande complémentaire au titre du CRPA ;
- Faire **figurer ces informations de façon claire, lisible et compréhensible** (problématiques de design).

AU REGARD DES BESOINS D'ACCOMPAGNEMENT ET DE PRÉCISIONS DES ADMINISTRATIONS, LA DÉFENSEURE DES DROITS RECOMMANDE AU GOUVERNEMENT :

- *D'assurer un soutien effectif aux administrations, collectivités et autres organismes concernés face au défaut d'expertise et à la faible incitation (en raison de l'absence de sanctions pour certaines obligations), et à l'enjeu de l'animation du dialogue entre les structures concernées, les usagers et plus globalement le débat public ;*
- *De déterminer quels systèmes algorithmiques et d'IA peuvent être utilisés par les services publics lorsqu'ils fondent une décision administrative individuelle partiellement automatisée de manière à garantir la conformité aux obligations du CRPA sur la transparence (précisément, celles relative à la communication des paramètres de traitement et, le cas échéant, de leur pondération, appliqués à la situation de l'intéressé) ;*
- *Concernant le guide des algorithmes publics proposé par Etalab¹⁷³, qui contient notamment des détails et un exemple de mention individuelle explicite :*
 - l'enrichir avec notamment des exemples d'informations complémentaires, assurer

une meilleure valorisation de ce guide et un accroissement des ressources d'Etalab ou d'une structure dédiée¹⁷⁴ pour ce faire ;

- ajouter des indications sur les « règles définissant les principaux traitements algorithmiques utilisés » ainsi que des exemples ;
 - l'adapter autant que de besoin pour les collectivités territoriales (notamment en termes d'exemples) et le faire porter par le ministère chargé des collectivités territoriales et l'Agence nationale de la cohésion des territoires (ANCT) ;
3. Une difficulté pour les administrations d'une part et pour les administrés d'autre part, liée au fait que la façon dont la demande d'information complémentaire faite par la personne concernée (par la décision administrative individuelle) est parfois formulée sans faire explicitement référence aux articles concernés du CRPA¹⁷⁵. Cela peut soulever pour les administrations des difficultés d'identification des termes de la demande qui leur est faite¹⁷⁶.

EN CONSÉQUENCE, LA DÉFENSEURE DES DROITS RECOMMANDE :

- **Aux administrations et autres organismes concernés, d'assurer la formation interne, dans les administrations et organismes concernés, afin que soient identifiées ces demandes d'informations supplémentaires par les agents les recevant, grâce à des exemples notamment ;**
- **Au Gouvernement et à destination des usagers, à partir des outils déjà développés par des associations¹⁷⁷, de :**
 - **Produire des modèles de demande** (d'informations complémentaires et de demande de publication du code source) ;
 - **Et de les rendre visibles et aisément accessibles sur les sites web des administrations et autres organismes concernés** (y compris de la Cada).

4. Enfin, des difficultés indéniables sont liées à la technicité des systèmes et au niveau de connaissance variable des personnes concernées : selon un sondage réalisé en 2023, seulement 24% des répondants déclarent savoir exactement ce que c'est que l'IA¹⁷⁸. Dans la lignée de la « fracture numérique »¹⁷⁹, on parle à ce sujet d'« opacité par analphabétisme technique »¹⁸⁰ qui se double d'une relative méconnaissance du droit à l'information et du droit d'accès. En ce sens, même si et quand les obligations légales sont respectées par les administrations concernées¹⁸¹, cela ne suffit pas à concrétiser l'exigence de transparence car les mentions et documents publiés restent sur le fond inaccessibles au public¹⁸², à qui revient la charge du travail de compréhension. Face à une décision administrative individuelle entièrement ou partiellement automatisée, les informations utiles sont celles qui facilitent sa compréhension et qui doivent permettre d'outiller l'utilisateur concerné ou le collectif qui les défend en permettant de contester la décision prise¹⁸³.

EN CONSÉQUENCE, LA DÉFENSEURE DES DROITS RECOMMANDE :

- **Au Gouvernement :**
 - Compte tenu des considérations du Conseil constitutionnel en 2018¹⁸⁴ et de ce que prévoit le règlement sur l'IA de l'UE en matière de droit à l'explication des décisions individuelles (mais seulement pour les décisions prises grâce à un système d'IA à haut risque), **de consacrer un « droit à l'explication » pour toutes les décisions administratives individuelles partiellement et entièrement automatisées**¹⁸⁵ ;
 - Compte-tenu de la technicité du sujet, de **soutenir la recherche dans ce domaine ainsi que les projets associatifs et collectifs visant à favoriser la compréhension et le débat public autour de ces sujets**, ce d'autant plus que les effets de ces systèmes se déploient aussi à l'échelle collective – en lien avec les risques de discrimination qui ne sont, dans des logiques algorithmiques, bien souvent identifiables qu'à cette échelle.

• **Aux administrations concernées :**

- De façon générale, et dans la lignée des recommandations du Conseil d'État¹⁸⁶, **d'associer les usagers du service public à tous les niveaux**, c'est-à-dire dès le choix du recours à un traitement algorithmique ou à un système d'IA, dans l'élaboration du libellé des informations, dans les forums mis en place le cas échéant au stade de l'évaluation du système et au sein des comités d'éthiques quand de tels comités existent ;
- **Pour que ce droit à l'explication trouve le plus possible à se concrétiser, s'emparer des outils existants et en développer d'autres.**

Pour que le système de décision administratif soit employé aussi pour expliquer, de façon personnalisée, le fondement algorithmique de la décision :

- Le système technique d'explication de la décision devrait être aligné sur le système technique de prise de cette décision et le système de prise de décision doit être adossé sur les éléments légaux qui fondent et justifient la décision¹⁸⁷ ;
- Évaluer la faisabilité d'une production automatique d'explications individualisées, détaillées et intelligibles par les parties prenantes du contrôle interne et externe du système de décision¹⁸⁸ ;
- À partir de l'outil Algocate¹⁸⁹, automatiser le volet « justifications ».

Pour faciliter l'appropriation, par les usagers concernés, des explications communiquées¹⁹⁰, alors que l'accès à l'algorithme ne suffit pas, à lui seul, à comprendre le résultat menant à une décision¹⁹¹ :

- Développer des outils d'explication présentant des contre-arguments, mettant en avant les limitations de la prédiction algorithmique ;
- Proposer des stratégies de jeux pour motiver les personnes à comprendre les informations qui leur sont présentées et à mieux chercher les informations pertinentes ;

- Mettre à dispositions des outils permettant à l'usager concerné de tester l'algorithme en changeant les données d'entrée, par exemple l'âge ou le quotient familial¹⁹².

Il apparaît que les efforts engagés pour **algorithmiser l'administration devraient avoir pour corollaire des efforts pour respecter les obligations prévues par la loi et le règlement¹⁹³, travailler sur l'explicabilité et mettre en œuvre les propositions doctrinales existantes** afin de permettre aux droits d'ores et déjà en vigueur pour les usagers du service public de se concrétiser et à mettre en œuvre un véritable droit à l'explication des décisions administratives individuelles entièrement ou partiellement automatisées.

C'est seulement à cette condition que la transparence pourra « *renouer avec sa dimension démocratique* » comme le souligne Jean-François Kerléo¹⁹⁴ en relevant que le processus de technicisation en cours dont elle est l'objet est caractérisé par « *un empilement de pratiques réservées de plus en plus à des initiés* ».

CONCLUSION

L'algorithmisation de l'administration, cinq ans après l'ouverture à la prise de décisions administratives individuelles entièrement automatisées, interroge le respect des droits des usagers à de multiples égards. Le Défenseur des droits s'est ici particulièrement intéressé aux décisions dites partiellement automatisées et à la substance de l'intervention humaine annoncée et effective, aux enjeux juridiques et pratiques de la transparence des systèmes et des décisions prises.

À l'heure où, dans l'administration, des systèmes automatisés de plus en plus complexes sont déployés ou sur le point de l'être, les réponses apportées à ces questions paraissent d'autant plus importantes qu'elles constituent, en fait, le socle sur la base duquel pourra se construire la confiance des usagers vis-à-vis du service public. Les recommandations de ce rapport se veulent ainsi une contribution pour renforcer les droits des usagers, assurer la qualité du service public et sa légitimité.

Pour s'y attacher, il est indispensable de (re) mettre les administrations au centre de l'expertise, en s'assurant que les compétences requises sont réunies en interne et que les équipes dialoguent en transversalité pour d'une part maîtriser les choix qui fondent le fonctionnement des systèmes comme les décisions individuelles qui en sont issues et d'autre part assurer une bonne appropriation de tous les enjeux que recèlent ces systèmes au sein du service public concerné.

Il apparaît tout aussi essentiel de soutenir la recherche en la matière, ainsi que l'émergence de collectifs et l'engagement d'associations qui s'emparent de ces sujets. La mobilisation de la société civile a déjà commencé, dans différents domaines notamment les systèmes d'IA déployés dans l'espace public via la vidéosurveillance algorithmique¹⁹⁵ ou ceux déployés dans le cadre de la lutte contre la fraude aux prestations sociales.

C'est enfin en analysant les résultats agrégés qu'il sera possible de questionner les objectifs poursuivis et de réorienter le système : seules des analyses portant sur les effets collectifs et structurels de ces systèmes peuvent utilement alimenter le débat public. En effet, face aux enjeux que soulèvent ces objets, la compréhension de la dimension collective de leurs effets apparaît seule de nature à révéler les atteintes qu'elles sont susceptibles d'entraîner sur les droits individuels des usagers et, in fine, à permettre de faire respecter ces droits.

NOTES

- ¹ Voir Défenseur des droits, Dématérialisation et inégalités d'accès aux services publics, 2019 ; Défenseur des droits, Dématérialisation des services publics : trois ans après, où en est-on, 2022.
- ² Défenseur des droits, Cnil, Algorithmes : prévenir l'automatisation des discriminations, 2020.
- ³ Défenseur des droits, Technologies biométriques : l'impératif respect des droits fondamentaux, 2021.
- ⁴ Voir Cnil, Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle, 2017.
- ⁵ Voir Conseil d'État, Intelligence artificielle et action publique : construire la confiance, servir la performance, 2022 ; Conseil d'État, Étude annuelle 2014, Le numérique et les droits fondamentaux, 2014.
- ⁶ Parmi une littérature nombreuse, on peut citer par exemple : Denis Merigoux, Marie Alauzen, Louis Gesbert et al., De la transparence à l'explicitabilité automatisée des algorithmes : comprendre les obstacles informatiques, juridiques et organisationnels, Rapport de recherche n°9535, INRIA, Paris, 2024, p.68 ; ENA, Promotion 2018-2019 MOLIÈRE, Éthique et responsabilité des algorithmes publics, Rapport collectif sur commande d'une administration centrale à la demande d'Étala, 2019.
- ⁷ Parmi une littérature particulièrement riche, on peut citer par exemple : Observatoire Data Publica, Justine Banuls (aut. principale), La transparence des algorithmes publics, Les cahiers de l'observatoire Data Publica, 2023.
- ⁸ Défenseur des droits, Rapport annuel d'activité 2023-2024.
- ⁹ *Ibid.*
- ¹⁰ Défenseur des droits, Dématérialisation des services publics : trois ans après, où en est-on ?, 2022.
- ¹¹ L. n° 78-17, 6 janvier 1978, relative à l'informatique, aux fichiers et aux libertés.
- ¹² Règl. (UE) n°2016/679 du Parlement européen et du Conseil, 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (règlement général sur la protection des données-RGPD).
- ¹³ L. n° 2016-1321, 7 octobre 2016, pour une République numérique, dite « loi Lemaire ».
- ¹⁴ Par exemple dans le domaine de l'éducation, pour ce qui concerne la procédure Parcoursup : voir l'article L. 612-3 du Code de l'éducation, tel que modifié par la loi n° 2018-166 du 8 mars 2018 relative à l'orientation et à la réussite des étudiants.
- ¹⁵ Règl. (UE) 2024/1689, 13 juin 2024, établissant des règles harmonisées concernant l'intelligence artificielle.
- ¹⁶ L'annexe III du Règlement sur l'intelligence artificielle liste, par type ou par domaine, les systèmes d'IA considérés comme à haut risque. Pour avoir une vue complète des systèmes entrant dans cette catégorie, il faut aussi se référer à l'annexe I du texte, conformément à l'article 6 du Règlement.
- ¹⁷ Convention-cadre sur l'intelligence artificielle et les droits de l'Homme, la démocratie et l'État de droit, adoptée par les ministres des affaires étrangères des 46 pays membres du Conseil de l'Europe, le 17 mai 2024 à Strasbourg.
- ¹⁸ Cnil, Cada, Etalab, Guide pratique de la publication en ligne et de la réutilisation des données publiques (open data), 2019, p. 26.
- ¹⁹ *Ibid.*
- ²⁰ Michel Blancard, Ouverture du code du calcul de l'impôt sur le revenu, années 2010 à 2015, sur <https://www.etalab.gouv.fr/>, publié le 17 septembre 2017.
- ²¹ Les algorithmes d'apprentissage automatique apprennent par entraînement : un algorithme reçoit d'abord des exemples dont les résultats sont connus, note la différence entre ses prévisions et les résultats corrects, et s'accorde avec des pondérations pour affiner la précision de ses prédictions jusqu'à ce qu'elles soient optimisées. En donnant aux ordinateurs « la capacité d'apprendre », la tâche complexe d'optimisation (le contrôle et l'analyse des variables dans les données disponibles pour en faire des prédictions précises sur l'avenir) revient donc à l'algorithme, auquel on peut même laisser le soin de préciser les caractéristiques à considérer en premier lieu.
- ²² On peut aussi dire que, selon la définition de Futura Science, « Le deep Learning s'appuie sur un réseau de neurones artificiels s'inspirant du cerveau humain. Ce réseau est composé de dizaines voire de centaines de « couches » de neurones, chacune recevant et interprétant les informations de la couche précédente. Le système apprendra par exemple à reconnaître les lettres avant de s'attaquer aux mots dans un texte, ou détermine s'il y a un visage sur une photo avant de découvrir de quelle personne il s'agit. »
- ²³ Règl. (UE) 2024/1689, 13 juin 2024, établissant des règles harmonisées concernant l'intelligence artificielle, art. 3.
- ²⁴ Cnil, Comment permettre à l'Homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle, 2017, p. 51.
- ²⁵ Boris Barraud, « L'algorithmisation de l'administration », Revue Lamy Droit de l'immatériel, 2018, n°150, pp. 42-54.
- ²⁶ Elisabeth Grosdhomme Lulin, Gouverner à l'ère du Big Data : promesses et périls de l'action publique algorithmique, Institut de l'entreprise, 2015, 78p. ; Gilles Drouet, Algorithmes et décisions publiques, CNRS, 2019.
- ²⁷ Voir Dinum, Feuille de route de la DINUM – Une stratégie numérique au service de l'efficacité de l'action publique

- sur <https://www.numerique.gouv.fr/>, publié le 9 mars 2023.
- ²⁸ Voir Etalab, Guide des algorithmes publics sur <https://etalab.github.io/algorithmes-publics/guide.html>.
- ²⁹ Voir en ce sens : Cour de cassation, Anonymisation des décisions : deux data scientists à la Cour sur <https://www.courdecassation.fr/>, publié le 15 janvier 2020 : « La Cour de cassation, en lien avec le ministère de la Justice, souhaite développer des techniques d'apprentissage automatique afin d'identifier les données à anonymiser dans les décisions de justice avant de les rendre accessibles et réutilisables. »
- ³⁰ Voir la loi n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions qui autorise, à titre expérimental, le recours à des algorithmes pour détecter plusieurs types d'événements, afin d'assurer la sécurité de manifestations sportives, récréatives ou culturelles qui, par l'ampleur de leur fréquentation ou par leurs circonstances, sont particulièrement exposées à des risques d'actes de terrorisme ou d'atteintes graves à la sécurité des personnes. Lesdits événements ont été prévus par le décret n° 2023-828 du 28 août 2023 et sont les suivants : présence d'objets abandonnés, présence ou utilisation de certaines armes, non-respect par une personne ou un véhicule du sens de circulation commun, franchissement ou présence d'une personne ou d'un véhicule dans une zone interdite ou sensible, présence d'une personne au sol à la suite d'une chute, mouvement de foule, densité trop importante de personnes, départs de feux. Ces événements peuvent être détectés grâce à l'utilisation de système d'IA appliqués aux images captées par les caméras de vidéo protection et les drones dans l'espace public.
- ³¹ Amélioration du diagnostic et de la prévention en matière médicale grâce à l'imagerie médicale et les algorithmes de modélisation des facteurs de risques. Le décret n° 2024-468 du 24 mai 2024 relatif à la mise en œuvre de l'espace numérique de santé (ENS) définit les conditions d'utilisation des données contenues dans l'ENS aux fins d'une prévention personnalisée de ses titulaires et ouvre la possibilité pour l'utilisateur de communiquer des questionnaires de santé et de recevoir des informations de prévention sur son ENS.
- ³² Voir en ce sens : Emile Marzolf, L'Éducation nationale renforce l'utilisation de l'IA dans toutes les classes du CP à la seconde, sur <https://acteurspublics.fr/>, publié le 23 janvier 2024. L'article précise que c'est le cas notamment de « Adaptatif'Maths », qui doit fournir aux enseignants « une assistance fondée sur un parcours adaptatif par groupes d'élèves, basé sur l'utilisation d'un algorithme de personnalisation dynamique » ou Lalilo qui est censé permettre aux élèves « grâce sa capacité de reconnaissance et de synthèse vocale, de progresser en compréhension orale et dans leur propre prononciation des mots. »
- ³³ Comme Nantes l'a essayé depuis 2020, la région Ile-de-France réfléchit en 2024 à mettre en œuvre un système d'anticipation du nombre de repas à livrer quotidiennement dans les lycées publics qu'elle gère (450 environ).
- ³⁴ Avant l'adoption de la loi n° 2016-1088 du 8 août 2016 relative au travail, à la modernisation du dialogue social et à la sécurisation des parcours professionnels, ses effets ont été évalués au moyen d'un logiciel (Worksim) qui modélise et simule le marché du travail français.
- ³⁵ Cnil, Intelligence artificielle : le plan d'action de la CNIL, sur <https://www.cnil.fr/fr>, publié le 16 mai 2023.
- ³⁶ Voir Info Gouv, IA : connaissez-vous Albert ? sur <https://www.info.gouv.fr/>, publié le 22 mai 2024.
- ³⁷ ENA, Promotion 2018-2019 MOLIÈRE, Rapport collectif sur commande d'une administration centrale [Etalab], Éthique et responsabilité des algorithmes publics, 2019, p. 1.
- ³⁸ Voir à cet égard Marianne Uguen, Le recours aux algorithmes dans l'action publique, Pr. Marcel Morabito (dir.), mémoire de master, sciences politiques, Sciences Po - Ecole d'affaires publiques, 2023, p. 12.
- ³⁹ Jacques Thyraud, Rapport fait au nom de la Commission des Lois constitutionnelles, de Législation, du Suffrage universel, du Règlement et d'Administration générale, sur le projet de loi adopté par l'Assemblée nationale, relatif à l'informatique et aux libertés, Sénat, 1977, p.22.
- ⁴⁰ Loi n°78-17, 6 janvier 1978, relative à l'informatique, aux fichiers et aux libertés, article 47 : « Aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de la personnalité de cette personne. »
- ⁴¹ Projet de loi n°490, 13 décembre 2017, relatif à la protection des données personnelles, p. 14.
- ⁴² L. n° 2018-493, 20 juin 2018, relative à la protection des données personnelles.
- ⁴³ Projet de loi relatif à la protection des données personnelles, Étude d'impact, 12 décembre 2017.
- ⁴⁴ Loi n°78-17, 6 janvier 1978, relative à l'informatique, aux fichiers et aux libertés, article 47 : « Aucune décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel, y compris le profilage (...) »
- ⁴⁵ Règl. (UE) n°2016/679, 27 avril 2016, RGPD, art. 22.
- ⁴⁶ L. n°2018-493, 20 juin 2018, relative à la protection des données personnelles, art. 21.
- ⁴⁷ Voir à cet égard Cnil, « 'Directive Police-Justice' : de quoi parle-t-on ? », 20 février 2019
- ⁴⁸ Cons. Const., 12 juin 2018, n° 2018-765 DC.
- ⁴⁹ Défenseur des droits, Cnil, Algorithmes : prévenir l'automatisation des discriminations, 2020.
- ⁵⁰ Nicolas Kayser-Bril, Austria's employment agency rolls out discriminatory algorithm, sees no problem, sur <https://algorithmwatch.org/en/>, publié le 6 octobre 2019.
- ⁵¹ Étude d'impact, Projet de loi relatif à la protection des données personnelles, 12 décembre 2017, p. 64.

- ⁵² CJUE, 1^{er} août 2022, n° C-184/20 ; CJUE, 5 juin 2023, n° C-204/21 ; CJUE, 4 juil. 2023, n° C-252/21 ; CJUE, 4 oct. 2024, n° C-21/23.
- ⁵³ Voir par exemple : Cnil, délibération n° 2023-134, 7 décembre 2023 portant avis sur projet de décret modifiant le décret n° 2003-485 du 5 juin 2003 relatif au recensement de la population : « La jurisprudence du Conseil d'État tend à considérer que la seule information selon laquelle une personne est affectée d'handicap, sans précision sur sa nature et son importance, ne serait pas une donnée de santé au sens du RGPD (CE, 10^{ème} et 9^{ème} sous-sections réunies, 19 juillet 2010, n°317182, Rec. ; CE, 10^e - 9^e sous-sections réunies, 28 mars 2014, n°361042 ; CE, 10^{ème} - 9^{ème} chambres réunies, 10 juin 2021, n°431875, Rec.). Ainsi, dans la mesure où l'indicateur "Gali" ne révélerait pas nécessairement un handicap et ne donnerait directement aucune information sur la nature ou la gravité d'un handicap éventuel, il pourrait être considéré qu'il n'est pas une donnée de santé au regard de la jurisprudence précitée. En tout état de cause, la Cnil portera une attention particulière, d'une part, à ce que la diffusion de cette information exclut toute réidentification d'individus ou stigmatisation de groupes de personnes vulnérables et, d'autre part, aux futures interconnexions et rapprochements de données qui pourraient entraîner une requalification des données recueillies au titre de l'indicateur "Gali" en données sensibles au sens du RGPD. »
- ⁵⁴ Selon la loi n°2008-496 du 27 mai 2008 portant diverses dispositions d'adaptation au droit communautaire dans le domaine de la lutte contre les discriminations, la liste des motifs protégés est la suivante : origine, sexe, situation de famille, grossesse, apparence physique, particulière vulnérabilité résultant de sa situation économique, apparente ou connue de son auteur, patronyme, lieu de résidence ou domiciliation bancaire, état de santé, perte d'autonomie, handicap, caractéristiques génétiques, mœurs, orientation sexuelle, identité de genre, âge, opinions politiques, activités syndicales, capacité à s'exprimer dans une langue autre que le français, appartenance ou non-appartenance, vraie ou supposée, à une ethnie, une nation, une prétendue race ou une religion déterminée.
- ⁵⁵ CJUE, 7 déc. 2023, n° C-634/21.
- ⁵⁶ Cf. La Quadrature du Net, Notation des allocataires : l'indécence des pratiques de la CAF désormais indéniable, sur <https://www.laquadrature.net/>, publié le 27 novembre 2023 ; Gabriel Geiger, Soizic Pénicaut, Manon Romain, Adrien Sénécet, Profilage et discriminations : enquête sur les dérives de l'algorithme des caisses d'allocations familiales, sur <https://www.lemonde.fr/>, publié le 4 décembre 2023.
- ⁵⁷ Cons. Const., 12 juin 2018, n° 2018-765 DC.
- ⁵⁸ Cnil, Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle, 2017.
- ⁵⁹ A., 17 oct. 2018, modifiant l'arrêté du 6 novembre 1996 modifié portant homologation des règles de répartition et d'attribution des greffons prélevés sur une personne décédée en vue de transplantation d'organes, art. 1^{er}, 2.2, NOR : SSAP1828392A.
- ⁶⁰ CRPA, art. L. 312-1-3 – pour les administrations de plus de 50 ETP et sous réserve des secrets protégés (cf. CRPA, article L. 311-5).
- ⁶¹ Soazig Le Nevé, À Sciences Po Bordeaux, l'algorithme maison favorise les boursiers et les élèves des lycées « cordées de la réussite », sur <https://www.lemonde.fr/>, publié le 23 mai 2023 .
- ⁶² En l'occurrence, l'arrêté modifié du 31 décembre 2020 portant création d'un traitement automatisé de données à caractère personnel dénommé « Parcoursup », NOR : ESR2035793A.
- ⁶³ Voir notamment Vincent Tiberj, Parcoursup ou la sélection par les algorithmes, sur <https://laviedesidees.fr/>, publié le 12 janvier 2021.
- ⁶⁴ Lucie Cluzel-Métayer, « Transparence et loyauté des algorithmes publics », in Jérémy Bousquet, Thibault Carrère, Sabrina Hammoudi, L'action publique algorithmique : risque et perspectives, Mare & Martin, Paris, 2023, pp. 167-180.
- ⁶⁵ On appelle « faux positif » la détection erronée d'une condition testée au cours d'un processus d'évaluation (par exemple, si le système d'IA identifie à tort un texte comme rédigé par un système d'IA, alors que le texte a été rédigé par un humain).
- ⁶⁶ On appelle « faux négatif » la non détection par le système d'IA d'une condition testée au cours d'un processus d'évaluation (par exemple, un système d'IA ne détecte pas le fait qu'une personne a chuté au sol alors que tel est le cas et que le système est censé identifier toutes les fois où cela arrive).
- ⁶⁷ L'article 3. 2 du règlement sur l'IA précise que le risque est entendu comme « la combinaison de la probabilité d'un préjudice et de la sévérité de celui-ci. »
- ⁶⁸ Le règlement sur l'IA de l'UE distingue (notamment) le fournisseur (l'organisme qui « développe ou fait développer un système d'IA ou un modèle d'IA à usage général et le met sur le marché ou met le système d'IA en service sous son propre nom ou sa propre marque, à titre onéreux ou gratuit » - article 3.3) et le déployeur (l'organisme « utilisant sous sa propre autorité un système d'IA » - article 3.4).
- ⁶⁹ Pour une vue du programme de travail du CEN/CLC/JTC 21 mis en place par le Comité européen de normalisation (CEN) et le Comité européen de normalisation électrotechnique (CENELEC), il est possible de consulter la page internet du CEN : https://standards.cenelec.eu/dyn/www/f?p=205:22:0:::FSP_ORG_ID,FSP_LANG_ID:2916257,25&cs=1827B89DA69577BF3631EE2B6070F207D.
- ⁷⁰ Règl. (UE) 2024/1689, 13 juin 2024, établissant des règles harmonisées concernant l'intelligence artificielle, art. 6.3.
- ⁷¹ Sur ce point, le règlement sur l'IA de l'UE renvoie au RGPD qui le définit comme suit (article 4.4) : « toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le

rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique ». Comme l'indique la Cnil « Un traitement de profilage repose sur l'établissement d'un profil individualisé, concernant une personne en particulier : il vise à évaluer certains de ses aspects personnels, en vue d'émettre un jugement ou de tirer des conclusions sur elle (...). Tout traitement de profilage doit faire l'objet d'une attention particulière, car il soulève par nature des risques importants pour les droits et libertés des personnes ». L'autorité donne l'exemple suivant : « Le profilage est au cœur du secteur des assurances, notamment lorsqu'il implique l'évaluation des caractéristiques du risque assurantiel pour en déterminer la fréquence, le coût moyen, le coût maximum du sinistre potentiel, la tarification et vérifier l'assurabilité du risque. Il n'est donc qu'un moyen, non une finalité » : Cnil, Droit des personnes et profilage : les spécificités du secteur de l'assurance, sur <https://www.cnil.fr/fr>, publié le 16 juillet 2021.

⁷² Cnil, Droit des personnes et profilage : les spécificités du secteur de l'assurance, sur <https://www.cnil.fr/fr>, publié le 16 juillet 2021

⁷³ Cf. les développements qui précèdent sur l'arrêt de la CJUE sur la notion de décision (CJUE, 7 déc. 2023, n° C-634/21).

⁷⁴ Tableau des principaux traitements algorithmiques du Département donnant lieu à une décision administrative individuelle, 25 mars 2024, publié sur <https://www.valdoise.fr>

⁷⁵ Conformément au dernier alinéa de l'article 47 de la loi « informatique et liberté » : « aucune décision par laquelle l'administration se prononce sur un recours administratif mentionné au titre Ier du livre IV du code des relations entre le public et l'administration ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel. »

⁷⁶ Groupe de travail « article 29 » sur la protection des données, Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679, 6 février 2018, p. 23.

⁷⁷ Conseil d'État, Étude annuelle 2014, Le numérique et les droits fondamentaux, La documentation française, 2014.

⁷⁸ Cnil, délib. n° 2016-180, 16 juin 2016.

⁷⁹ Cnil, délib. n° 2023-058, 8 juin 2023.

⁸⁰ Cons. Const., 27 déc. 2019, n° 2019-796 DC.

⁸¹ Groupe de travail « Article 29 » sur la protection des données, *op. cit.* p. 23.

⁸² Voir notamment Liane Huttner, La décision de l'algorithme, étude de droit privé sur les relations entre l'humain et la machine, Dalloz, 2024.

⁸³ Voir notamment Winston Maxwell, Le contrôle humain des systèmes algorithmiques – un regard critique sur l'exigence d'un « humain dans la boucle », mémoire original pour présenter l'habilitation à diriger des recherches, droit, Université Paris 1 Panthéon-Sorbonne, 2022.

⁸⁴ Conseil de l'Europe, Comité d'experts sur les dimensions des droits de l'Homme dans le traitement automatisé

des données et les différentes formes d'intelligence artificielle (MSI-AUT), Étude [DGI(2019)05] sur les incidences des technologies numériques avancées (dont l'intelligence artificielle) sur la notion de responsabilité, sous l'angle des droits humains, 2019, p. 68 ; Madeleine Clare Elish, Letting Autopilots Off the Hook: Why do we blame humans when automation fails?, sur <https://slate.com/>, publié le 16 juin 2016.

⁸⁵ C. éduc., art. D. 612-1-13.

⁸⁶ Ce que confirme la note de cadrage des ministères de l'Éducation nationale et de la Jeunesse, des Sports et de l'Enseignement supérieur et de la Recherche, L'Utilisation du module d'aide à la décision Parcoursup : les obligations du RGPD, 2021, p. 3 : « Dans la mesure où la décision n'est pas entièrement automatisée, les exigences du RGPD en matière de traitement algorithmique, qui impliquent notamment d'informer les personnes concernées de la logique sous-jacente de l'algorithme et de l'importance et des conséquences de cet algorithme, ne sont pas applicables à ces traitements. »

⁸⁷ Cons. Const., 3 avril 2020, n° 2020-824 QPC, §14.

⁸⁸ Liane Huttner, *op. cit.*

⁸⁹ Anne Debet, « Parcoursup : le Conseil constitutionnel consacre un droit constitutionnel d'accès aux documents administratifs », Communication – Commerce électronique, n°9, 2020, comm. 65.

⁹⁰ Voir Jacques Grosperin, « Algorithmes locaux dans Parcoursup – Communication », in. Comptes rendus de la commission de la culture, de l'éducation et de la communication [Sénat], 2019 : à propos de l'université Paris Descartes « Une fois le pré-classement opéré, les opérations humaines ont débuté : la commission a regardé tous les dossiers des candidats du secteur – environ 600 dossiers – et les 100 premiers hors secteur (...) »

⁹¹ Société informatique de France (SIF), Bacheliers : le Parcoursup du combattant, 2024.

⁹² A., 17 juillet 2017, portant création par le ministère de l'éducation nationale d'un traitement automatisé de données à caractère personnel dénommé « Affelnet lycée », NOR : MENE1719988A.

⁹³ Défenseur des droits, déc. n° 2023-140, 26 juin 2023.

⁹⁴ Liane Huttner, *op. cit.*

⁹⁵ À cet égard, voir notamment Nicolas Spatola, La supervision humaine reste essentielle pour interpréter les résultats de l'IA et prendre des décisions éclairées, sur <https://acteurspublics.fr/>, publié le 17 juillet 2024.

⁹⁶ Voir Conseil d'État, Intelligence artificielle et action publique : construire la confiance, servir la performance, 2022.

⁹⁷ Voir Conseil d'État, Intelligence artificielle et action publique : construire la confiance, servir la performance, 2022.

⁹⁸ Conseil d'État, Étude annuelle 2014, Le numérique et les droits fondamentaux, La documentation française, 2014, prop. 23.

⁹⁹ Winston Maxwell, *op. cit.*

- ¹⁰⁰ Applicabilité de l'article 22 du RGPD, repris à l'article 47 de la loi « Informatique et Libertés » pour les traitements relevant du RGPD.
- ¹⁰¹ ENA, Promotion 2018-2019 MOLIERE, *op. cit.*
- ¹⁰² Dans le cas d'Affelnet, la fiche barème laissant apparaître la seule valeur de « 0 » dans la catégorie relative aux évaluations scolaires aurait ainsi fait l'objet d'une vérification en prenant en compte les résultats scolaires de l'élève.
- ¹⁰³ Voir notamment Winston Maxwell, *op. cit.*
- ¹⁰⁴ Conseil de l'Europe, Comité d'experts sur les dimensions des droits de l'Homme dans le traitement automatisé des données et les différentes formes d'intelligence artificielle (MSI-AUT), *op. cit.*
- ¹⁰⁵ Winston Maxwell, *op.cit.*
- ¹⁰⁶ CRPA, art. R. 311-3-1-2.
- ¹⁰⁷ Liane Huttner, *op. cit.*
- ¹⁰⁸ Jean-Marc Sauvé, *Transparence, valeurs de l'action publique et intérêt général, Exposé d'ouverture lors du colloque « Culture du secret contre transparence sans limite : quel équilibre pour garantir l'intérêt général ? »*, Transparence International France, Assemblée nationale, 2011.
- ¹⁰⁹ Cf. l'article 41 de la Charte des droits fondamentaux de l'Union européenne, 2000/C 364/01, 2000, qui consacre le « droit à une bonne administration » : « 1. Toute personne a le droit de voir ses affaires traitées impartialement, équitablement et dans un délai raisonnable par les institutions, organes et organismes de l'Union. 2. Ce droit comporte notamment : a) le droit de toute personne d'être entendue avant qu'une mesure individuelle qui l'affecterait défavorablement ne soit prise à son encontre ; b) le droit d'accès de toute personne au dossier qui la concerne, dans le respect des intérêts légitimes de la confidentialité et du secret professionnel et des affaires ; c) l'obligation pour l'administration de motiver ses décisions. [...] »
- ¹¹⁰ L'article 47, 2 de la loi « Informatique et Libertés » prévoit en effet que les décisions concernées comportent cette mention explicite « à peine de nullité ».
- ¹¹¹ Pour une approche critique du concept de transparence, voir par exemple : Louis Vuarin, Véronique Steyer. *L'injonction à la transparence : un levier réglementaire à double tranchant pour les organisations*, Bulletin de l'Association Française pour l'Intelligence Artificielle, 2023, pp.26-35.
- ¹¹² Jean-François Kerléo, *La transparence s'est muée en une politique publique*, sur <https://acteurspublics.fr/>, publié le 22 avril 2024.
- ¹¹³ Selon l'article L. 211-5 du CRPA : « La motivation exigée par le présent chapitre doit être écrite et comporter l'énoncé des considérations de droit et de fait qui constituent le fondement de la décision. »
- ¹¹⁴ Lucie Cluzel-Métayer, *op. cit.*
- ¹¹⁵ Félicien Vallet, *Entretien avec Clément Henin, Daniel Le Métayer : « Fournir des explications du fonctionnement des algorithmes compréhensibles par des profanes »*, sur <https://linc.cnil.fr/>, publié le 6 janvier 2021.
- ¹¹⁶ En prenant l'exemple d'un système d'aide à la décision pour traiter des demandes de prêt bancaire, Clément Héning et Daniel Le Métayer indiquent qu'une explication pourrait être « votre demande de prêt est refusée car votre taux d'endettement dépasserait le tiers de vos revenus ». Cette information explique le refus mais ne le justifie pas. La justification pourrait être : « votre demande de prêt est refusée car la loi interdit aux banques d'octroyer des prêts conduisant à un taux d'endettement dépassant le tiers des revenus », *Ibid.*
- ¹¹⁷ Voir notamment les actions et réflexions engagées par le collectif « Le mouton numérique » et l'association « La Quadrature du Net ».
- ¹¹⁸ En ce sens, voir Simon Chignard, *L'indispensable transparence des algorithmes publics*, sur <https://dataanalyticspost.com/>, publié le 12 juin 2019.
- ¹¹⁹ Soizic Pénicaud, *Simon Chignard, Appliquer le principe du Gouvernement Ouvert aux algorithmes publics, Webinaire*, sur <https://www.modernisation.gouv.fr/> [Dinum], enregistré le 31 mai 2021.
- ¹²⁰ En ce sens, voir Défenseur des droits, Cnil, *Rapport - Algorithmes : prévenir l'automatisation des discriminations*, 2020.
- ¹²¹ En ce sens, voir Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, *op. cit.* p. 68.
- ¹²² En ce sens, voir Cada, « Entretien avec Fabien Tarissan », in *Rapport d'activité 2022-2023, 2024*, pp. 44-47.
- ¹²³ Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, *op. cit.*, p. 7. - l'exemple donné dans ce rapport est celui de la Cnaf qui dans la réponse adressée au collectif *Changer de cap* indique 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 information. »
- ¹²⁴ Voir l'article 71-1 de la Constitution du 4 octobre 1958 et l'article 4, 1° de la loi organique n° 2011-333 du 29 mars 2011 relative au Défenseur des droits.
- ¹²⁵ Liane Huttner, *op. cit.*
- ¹²⁶ Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, *op. cit.*, p. 9. - L'obfuscation peut être définie comme « 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. »
- ¹²⁷ Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, *op. cit.*
- ¹²⁸ Règl. (UE) 2024/1689, 13 juin 2024, établissant des règles harmonisées concernant l'intelligence artificielle, cons. 59.
- ¹²⁹ Cnil, *Explicabilité (IA)*, sur <https://www.cnil.fr/>.
- ¹³⁰ Nicolas Maudet, Grégory Bonnet, Gaël Lejeune et al., *IA & explicabilité*, Bulletin de l'Association française pour l'Intelligence Artificielle, n°116, 2022.
- ¹³¹ Voir à cet égard le projet lancé à l'été 2024 par la Cnil et son laboratoire d'innovation numérique (Linc),

sur une base de données contenant environ 12 000 publications : Romain Pialat, Lancement d'un projet sur l'explicabilité dans le domaine de l'intelligence artificielle, sur <https://linc.cnil.fr/>, publié le 1 août 2024.

¹³² Félicien Vallet, *op. cit.*

¹³³ *Ibid.*

¹³⁴ « Ce domaine souffre encore beaucoup de ce que le chercheur Tim Miller et ses collègues ont appelé le syndrome des détenus qui veulent diriger l'asile (« beware of inmates running the asylum »), en l'occurrence des experts d'IA expliquant leurs propres productions. », *ibid.*

¹³⁵ À cet égard, il peut être renvoyé aux travaux mettant en lumière la substitution massive de l'explicabilité par la mise en avant d'indicateurs de performance. Voir Louis Vuarin, Véronique Steyer, « Le principe d'explicabilité de l'IA et son application dans les organisations », in Réseaux : communication, technologie, société, N° 240 (4), pp. 179-210.

¹³⁶ Cnil, Conformité RGPD : comment informer les personnes et assurer la transparence ?, sur <https://www.cnil.fr/>, publié le 29 juillet 2019.

¹³⁷ Cf. Cada, Le rôle de la Cada, sur <https://www.cada.fr/>.

¹³⁸ Cf. Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679, qui indiquent qu'il est « de bonne pratique de fournir les informations » même quand la prise de décision et le profilage automatisés ne satisfont pas aux critères de l'article 22 du RGPD (p. 28).

¹³⁹ Conformément aux articles L. 311-3-1 et R. 311-3-1-1 du CRPA.

¹⁴⁰ Secrets listés à l'article L. 311-5, 2° du CRPA.

¹⁴¹ Conformément à l'article R. 311-3-1-2 du CRPA.

¹⁴² C'est-à-dire l'État, comme le précise l'article L. 300-2 du CRPA, les collectivités territoriales, les autres personnes de droit public ou les personnes de droit privé chargées d'une mission de service public.

¹⁴³ Conformément à l'article D. 312-1-4 du CRPA.

¹⁴⁴ Secrets listés à l'article L. 311-5, 2° du CRPA.

¹⁴⁵ Cada, Cnil, Guide pratique de la publication en ligne et de la réutilisation des données publiques (« open data »), 2016, p. 26 : « Pour qu'un algorithme puisse être mis en œuvre par un ordinateur, il faut qu'il soit exprimé dans un langage informatique, transcrit en un programme (texte composé de commandes écrites, également appelé 'code source'. »

¹⁴⁶ CRPA, art. L. 300-2.

¹⁴⁷ Une demande d'accès aux documents administratifs, dite aussi « demande Cada », permet à toute personne de demander par oral ou par écrit l'accès à un document administratif. La page suivante du site service-public explique la procédure à suivre : <https://www.service-public.fr/particuliers/vosdroits/F2467>.

¹⁴⁸ CRPA, art. L. 311-9.

¹⁴⁹ Comme le précise l'article L. 312-1-1 du CRPA, ces administrations doivent publier en ligne les documents

qu'elles communiquent, sous réserve des secrets protégés.

¹⁵⁰ CRPA, art. D. 312-1-1-1.

¹⁵¹ Règl. (UE), 2016/679, 27 avril 2016, RGPD, art. 15.

¹⁵² Règl. (UE), 2016/679, 27 avril 2016, RGPD, art. 30.

¹⁵³ Cons. Const., 12 juin 2018, n° 2018-765 DC.

¹⁵⁴ CRPA, art. R. 311-3-1-2.

¹⁵⁵ Source : Cnil, Paramètre (IA), sur <https://www.cnil.fr/>.

¹⁵⁶ Winston Maxwell, *op. cit.*

¹⁵⁷ À l'article 86 du règlement sur l'IA de l'UE, il est précisé que ce droit ne s'applique que dans la mesure où il n'est pas prévu par ailleurs en droit de l'Union européenne et que ce droit s'applique sauf exceptions prévues dans le droit de l'UE ou dans le droit national.

¹⁵⁸ Conformément à l'article L. 311-5, 2° du CRPA.

¹⁵⁹ À noter que pour certaines autorités, comme le Défenseur des droits, seuls sont opposables le secret de la défense nationale, de la sûreté de l'État et de la politique extérieure (article 20 de la loi organique n° 2011-333 du 29 mars 2011 – voir à cet égard : Cnil, Recueil des procédures des tiers autorisés, 2020.)

¹⁶⁰ Voir Cada, Rapport d'activité 2022-2023, 2024, p. 42, à propos de son avis n°20213847 du 13 janvier 2022.

¹⁶¹ Cada, avis n°20226179, 15 déc. 2022.

¹⁶² Selon la décision du Conseil constitutionnel, 3 avril 2020, n°2020-834 QPC : « Est ainsi écartée, d'une part, l'application de l'article L. 311-3-1 dudit code, qui imposerait à l'administration, sous réserve des secrets protégés par la loi, de communiquer au candidat qui en fait la demande les règles définissant le traitement algorithmique et les principales caractéristiques de sa mise en œuvre. Est également exclue, d'autre part, l'application de l'article L. 312-1-3 du même code, qui obligerait l'administration, sous la même réserve, à publier en ligne les règles définissant les principaux traitements algorithmiques utilisés dans l'accomplissement de ses missions ».

¹⁶³ CRPA, art. L. 311-6.

¹⁶⁴ Cada, *op. cit.*, p. 43 : à propos de son avis n°20230314 du 30 mars 2023.

¹⁶⁵ CE, 8 nov. 2017, n° 375704.

¹⁶⁶ Conformément à l'article L. 300-2 du CRPA, sont concernés l'État, les collectivités territoriales, les autres personnes de droit public mais aussi les personnes de droit privé chargées d'une mission de service public de plus de 50 ETP (article D. 312-1-4 du CRPA).

¹⁶⁷ Le Conseil d'état dans son rapport précité sur l'IA indique à propos de cette obligation que « son effectivité est douteuse en pratique » (p. 120) ; Lucie Cluzel-Métayer, *op. cit.*, indique que « rares sont les administrations qui ont initié ce délicat travail de cartographie de leurs traitements » ; Emile Marzolf, Les algorithmes, nœud gordien de la transparence administrative, sur <https://acteurspublics.fr/>, publié le 17 avril 2024 : « Dans les faits, la transparence algorithmique n'est que très peu mise en œuvre. »

- ¹⁶⁸ Emile Marzolf, Le département d'Ille-et-Vilaine fait toute la lumière sur ses algorithmes, sur <https://acteurspublics.fr/>, publié le 24 juillet 2024.
- ¹⁶⁹ À cet égard, il est possible de citer notamment : France Travail, Algorithmes, Tout savoir sur les algorithmes publiés par France Travail, sur <https://www.francetravail.fr/> ; Ille & Vilaine, Recensement des traitements algorithmiques, sur <https://data.ille-et-vilaine.fr/> ; Val-d'Oise le département, Tableau des principaux traitements algorithmiques du Département donnant lieu à une décision administrative individuelle, 2024.
- ¹⁷⁰ Cf. le site de l'ouvre-boîte, association dont l'objet est d'obtenir la publication effective des documents administratifs : <https://ouvre-boite.org/>.
- ¹⁷¹ Cf. le site de l'observatoire des algorithmes publics (ODAP) qui crée et rassemble des informations sur les algorithmes utilisés par les administrations françaises : <https://odap.fr/>.
- ¹⁷² Voir à cet égard Emile Marzolf, Les responsables de l'accès aux documents administratifs, une courroie de transmission en mal de notoriété, sur <https://acteurspublics.fr/>, publié le 17 avril 2024.
- ¹⁷³ Le Guide est disponible en ligne, via ce lien : <https://guides.etalab.gouv.fr/algorithmes/mention/#dans-quels-cas-l-obligation-de-mention-explicite-s-applique-t-elle>.
- ¹⁷⁴ À cet égard, la feuille de route de la Dinum ne fait pas état du sujet de la transparence des algorithmes publics : Dinum, Une stratégie numérique au service de l'efficacité de l'action publique, Feuille de route de la DINUM, 2023.
- ¹⁷⁵ CRPA, art. R. 311-3-1-1 et R. 311-3-1-2.
- ¹⁷⁶ C'était par exemple le cas dans les faits ayant donné lieu à la décision du Défenseur des droits n° 2023-140 du 26 juin 2023.
- ¹⁷⁷ Voir notamment la page relatif au processus de demande créée par l'association Ouvre-boîte et l'initiative Ma dada de l'association OpenKnowledge France.
- ¹⁷⁸ <https://www.ipsos.com/fr-fr/51-des-francais-ne-se-declarent-pas-particulierement-effrayes-par-lintelligence-artificielle>
- ¹⁷⁹ Markus Krajewski « Source Code Criticism : On Programming as a Cultural Technique and its Judicial Linkages », in. *Journal of Cross-disciplinary Research in Computational Law* 1.3, 2023.
- ¹⁸⁰ Jenna Burrell, « How the machine 'thinks' : Understanding opacity in machine learning algorithms », in. *Big Data & Society* 3.1, 2016.
- ¹⁸¹ Voir en ce sens : Dinum, Plan d'action logiciels libres et communs numériques, sur <https://www.numerique.gouv.fr/>, publié le 10 novembre 2021.
- ¹⁸² En ce sens, voir : Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, op. cit. : « la publication en source ouverte est une condition nécessaire, mais non suffisante de la transparence de l'administration vis-à-vis du public » ; « toutes les décisions prises sur le fondement d'un traitement automatisé ne font pas l'objet d'explication et lorsque le responsable fournit des explications, elles se caractérisent par leur brièveté voire un certain hermétisme » ; « la mise à disposition sur internet des documents est loin de régler la question. »
- ¹⁸³ Winston Maxwell, *op. cit.*
- ¹⁸⁴ Cons. Const., 12 juin 2018, n° 2018-765 DC.
- ¹⁸⁵ Proposition faite notamment par Judith Rochfeld, « Données à caractère personnel - Droit de ne pas subir une décision fondée sur un traitement automatisé », in. *IP/IT et Communication*, 2020, p. 23.
- ¹⁸⁶ Voir Conseil d'État, Intelligence artificielle et action publique : construire la confiance, servir la performance, 2022.
- ¹⁸⁷ Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, *op. cit.*
- ¹⁸⁸ À partir du projet Catala développé par l'INRIA (méthodologie de programmation qui relie le code source du système aux textes qui fondent sa décision automatisée)
- ¹⁸⁹ Outil développé par Clément Hénin et Daniel Le Métayer. Voir par exemple Félicien Vallet, *op. cit.*
- ¹⁹⁰ Ces 3 pistes sont proposées par Winston Maxwell, *op. cit.*
- ¹⁹¹ Cada, « Entretien avec Fabien Tarissan », in Rapport d'activité 2022-2023, 2024, pp. 44-47.
- ¹⁹² En ce sens, voir notamment Lucas Boncourt, Le défi de l'explicabilité des algorithmes publics, sur <https://www.banquedesterritoires.fr/>, publié le 17 septembre 2019.
- ¹⁹³ Ceci à l'heure où certains acteurs font le constat que l'obligation de publication des règles définissant les principaux traitements algorithmiques utilisés par les administrations dans l'accomplissement de leurs missions lorsqu'ils fondent des décisions individuelles est restée « lettre morte » depuis 2016. Voir en ce sens : Emile Marzolf, Le département d'Ille-et-Vilaine fait toute la lumière sur ses algorithmes, sur <https://acteurspublics.fr/>, publié le 24 juillet 2024.
- ¹⁹⁴ Jean-François Kerléo, La transparence s'est muée en une politique publique, sur <https://acteurspublics.fr/>, publié le 22 avril 2024.
- ¹⁹⁵ Cf. La Quadrature du Net, VSA, La surveillance bâtit son empire, sur <https://www.laquadrature.net/>.

Défenseur des droits - TSA 90716 - 75334 Paris Cedex 07 - 09 69 39 00 00

defenseurdesdroits.fr

