HAL open science

# La part de l'Homme et celle de la machine dans les décisions " automatisées ". Propositions pour une réécriture de l'article 22 du RGPD

Emmanuel Netter

▶ **To cite this version:**

Emmanuel Netter. La part de l'Homme et celle de la machine dans les décisions " automatisées ". Propositions pour une réécriture de l'article 22 du RGPD. Vanessa Barbé. Les algorithmes et le droit, Mare et Martin, inPress. hal-03555156

**HAL Id: hal-03555156**

**https://hal.science/hal-03555156v1**

Submitted on 3 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# La part de l'Homme et celle de la machine dans les décisions « automatisées ». Propositions pour une réécriture de l'article 22 du RGPD.

Emmanuel Netter
*Professeur de droit privé à l'université d'Avignon*

> *« La décision que nous avons rendue est susceptible de faire l'objet de critiques, chacun a le droit d'avoir un sentiment. Confondre ce sentiment avec un jugement, c'est autre chose. Pour arriver à ce jugement, le tribunal a lu 4 800 pages de procédure, consacré vingt-six heures aux débats, écouté 15 avocats et un procureur, et s'est réuni en collégialité pendant plusieurs heures. Ce n'est pas pour ça que notre décision est la bonne, mais ce qui fait que cette décision est un jugement, c'est que nous avons, chacun, tenu à distance nos émotions, nos impulsions ».*

**1. Bien décider.** Qu'est-ce qu'une bonne décision ? La réponse ci-dessus reproduite est celle de Michaël Humbert, président du tribunal correctionnel ayant statué dans l'affaire « Mila », du nom de cette jeune femme qui avait fait l'objet de vives attaques sur les réseaux sociaux en raison de ses positions relatives à l'islam[1]. Le type de décision dont il est ici question est donc l'un des plus solennels, l'un des plus lourds d'enjeux qui se puisse concevoir : faut-il précipiter sur un prévenu les foudres de la justice pénale, et avec quelle puissance ? Avant de parvenir à son jugement, le Tribunal a entendu les avocats et représentants du ministère public, échangé avec eux, débattu en son sein. Des discussions, du temps et des doutes : autant d'éléments qui semblent caractéristiques des processus décisionnels humains.

**2. Décider avec la machine.** L'Estonie avait pourtant annoncé il y a deux ans, avec le sens du marketing qui la caractérise s'agissant des technologies numériques, qu'elle comptait déléguer à des systèmes informatiques le soin de rendre seuls des décisions en matière pénale[2]. Encore était-il précisé que tout prévenu insatisfait aurait le droit à ce qu'un juge humain reprenne l'affaire dans son entier. Il suffit d'imaginer, ce que l'Estonie ne précisait pas dans sa communication au grand public, que le parquet disposait sans doute d'un droit symétrique en cas de décision trop clémente, pour dégonfler quelque peu l'innovation et rejoindre le constat d'Antoine Garapon selon lequel il y a actuellement une part de « bluff » autour des systèmes de prise de décision automatisés[3].

Florence G'Sell invitait toutefois à ne pas sous-estimer l'appétit des justiciables pour un système judiciaire « plus rapide et surtout plus simple »[4]. Le traitement informatisé se caractérise, il est vrai, par sa célérité, mais ce n'est pas son seul attrait. Revenons aux qualités présentées par M. Humbert comme concourant à une décision de qualité. Parmi elles, figure la capacité à tenir compte de manière exhaustive de milliers de pages de procédure, de dizaines d'heures de débat, une performance située à la limite des capacités de l'esprit humain. Multipliez cependant ce volume d'informations par dix, cent ou mille : la machine ne rencontrera aucune difficulté. Quant à l'aptitude, évoquée par le président du tribunal, à « tenir à distance [ses] émotions, [ses] impulsions », elle est à l'évidence un attribut naturel des systèmes de calcul.

1 H. Seckel, « "Ce que vous ne faites pas dans la rue, ne le faites pas sur les réseaux sociaux" : de quatre à six mois avec sursis et une leçon de conduite pour les harceleurs de Mila », article lemonde.fr du 7 juillet 2021.
2 V. par ex. H. Gardette, « Les robots feront-ils de bons juges ? », émission France Culture du 23 mai 2019 avec F. G'Sell, B. Jean et E. Polnas.
3 « Méfiance sur les "robots-juges" dans les tribunaux », article 24heures.ch du 26 avril 2019.
4 Émission précitée « Les robots feront-ils de bons juges ? ».

Les ordinateurs disposent donc d'un certain nombre d'atouts qui peuvent être utiles à une bonne prise de décision[5]. Il n'y rien d'étonnant à cela : décider, c'est s'engager dans l'une des voies ouvertes, en abandonnant les autres, sur la base des informations disponibles ; or, l'informatique est la science du traitement de l'information.

**3. Qui tient l'outil ?**. Encore faut-il que les capacités des machines soient employées à bon escient. L'ouvrage célèbre de Cathy O'Neil, *Weapons of Math Destruction* a largement contribué à attirer l'attention sur des mésusages des traitements algorithmiques décisionnels[6]. L'auteur relate notamment l'histoire d'une enseignante américaine, licenciée en dépit de l'excellente opinion qu'avaient d'elle élèves, parents et collègues. « L'algorithme » lui avait en effet attribué un score la plaçant parmi les professeurs les plus médiocres de sa région. La manière dont l'indicateur était construit n'était pas révélée ; le destinataire de la décision n'était pas en mesure de présenter ses observations. Un processus décisionnel opaque et arbitraire devenait incontestable parce qu'il avait été repeint aux couleurs de la science. L'objectivité était présentée comme certaine ; l'équité du résultat, présumée de manière irréfragable.

L'histoire racontée par Cathy O'Neil est certes édifiante et terrifiante. Mais elle résulte bien d'un malentendu fondamental : il n'y a pas de raison de se prosterner devant l'ordinateur comme devant une divinité païenne. Cela revient à vénérer un tournevis ou une paire de ciseaux. Demandons-nous plutôt qui mobilise ces outils, et ce qu'il cherche à faire. L'irruption de l'informatique dans un processus de décision n'est ni une bonne ni une mauvaise nouvelle en soi. Tout dépend de la façon dont elle est utilisée, et de la manière dont elle s'articule avec la volonté humaine.

**4. L'ordinateur ne décide jamais seul**. Toute prétendue « décision informatique » encapsule en effet une décision humaine. Plutôt que de les opposer l'une à l'autre comme si elles étaient étrangères, il est donc préférable d'observer comment elles s'interpénètrent.

Considérons une première famille de traitements informatiques, que nous qualifierons de « déductifs »[7]. L'humain fournit à la machine un ensemble de règles prêtes à l'emploi, qu'il ne reste qu'à exécuter. À ce stade, rien ne permet de préjuger favorablement ni défavorablement du contenu de ces règles. Afin de déterminer la performance d'un ensemble de salariés, on peut imaginer attribuer à chaque lettre de leur nom de famille un nombre correspondant à son rang dans l'ordre alphabétique, d'additionner l'ensemble, de le diviser par le numéro de leur département de naissance et enfin de le multiplier par la température actuelle en degrés Celsius. Le score en résultant aura beau être parfaitement absurde, si on ne permet à personne d'accéder à la manière dont il est construit, il sera difficilement contestable. Mais on peut imaginer un autre exemple. Pour l'épreuve du 100 mètres des Jeux olympiques, je peux demander à une machine de lancer l'épreuve par un signal sonore, d'arrêter le chronomètre de chaque coureur au moment où il sera détecté qu'il coupe la ligne d'arrivée, puis d'afficher le résultat sur le tableau du stade par ordre croissant de temps de course. Le processus est éventuellement à la merci d'incidents techniques, mais il est parfaitement sain dans sa conception générale. Retenons que, dans cette première famille d'automatismes, l'humain fixe directement et *a priori* l'ensemble des règles applicables.

---

5 Sur cette question et l'ensemble du thème de cette étude : S. Desmoulin-Canselier et D. Le Métayer, Décider avec les algorithmes. Quelle place pour l'Homme, quelle place pour le droit ?, Dalloz, 2020.

6 C. O' Neil, Weapons of Math Destruction, Crown, 2016. L'ouvrage a été traduit en français sous le nom d'Algorithmes : la bombe à retardement, Les Arènes, 2018.

7 Pour une présentation des différents types de traitements algorithmiques, y compris ceux que nous qualifions plus loin « d'inductifs », V. par ex. S. Fischman et B. Gomes, « Intelligences artificielles et droit du travail : contribution à l'étude du fonctionnement des plateformes numériques », in Intelligence artificielle, gestion algorithmique du personnel et droit du travail, dir. P. Adam, M. Le Friant et Y. Tarascewicz, Thèmes et commentaires, Dalloz, 2020, spéc. I.

Il en va autrement dans la famille des traitements que nous qualifierons « d'inductifs », en ce qu'ils sont supposés examiner des cas concrets, des données d'expérimentation, puis en tirer eux-mêmes les règles qu'il convient d'appliquer afin d'aboutir à un résultat satisfaisant, ce que l'on appelle « apprentissage automatique » ou « apprentissage machine »[8]. C'est ainsi que fonctionne notamment « l'apprentissage supervisé » : l'être humain fournit à l'ordinateur un corpus d'apprentissage, constitué de données d'entraînement qu'il a lui-même rattachées à une catégorie. Voici 100 000 photos dont j'indique qu'elles représentent des chats, autant d'images de chiens, à toi de découvrir et de formuler les critères pertinents pour différencier ces deux animaux. Il pourrait également s'agir de *curriculum vitae* de candidats à l'emploi dont la carrière a ensuite été un franc succès ou un échec patent, ou de profils d'emprunteurs bancaires ayant honoré leurs échéances pour les premiers, fait défaut pour les seconds. L'apprentissage machine peut encore prendre une forme non supervisée, mais guidée par « renforcement ». L'être humain intervient ici — on l'a dit, il intervient *toujours* — par l'assignation d'un objectif. Il peut s'agir de faire gravir des marches à un robot qui ne sait pas encore se déplacer, d'apprendre à devenir le meilleur joueur de Go du monde ou de minimiser le temps que des employés travaillant dans un gratte-ciel passent à attendre les ascenseurs. Le système va expérimenter différentes approches, et pourra mesurer lui-même leur performance au regard de l'objectif initial.

**5. Les « boîtes noires ».** Il résulte de ce qui précède que les systèmes informatiques visant à la prise de décisions ne sont intrinsèquement ni objectifs ni arbitraires, mais qu'ils sont des outils maniés par la volonté humaine. Il apparaît par ailleurs que les logiques qui gouvernent ces algorithmes peuvent rester inaccessibles aux destinataires des décisions. Dans le cas des traitements « déductifs », cette opacité résultera toujours d'un choix du responsable de traitement puisque, par définition même, les règles appliquées y sont connues *ex ante*. Il est donc en mesure de les dévoiler, mais il peut être tenté de les encapsuler dans un procédé technique qui ne les laisse pas apparaître, seule la décision étant révélée. En revanche, certains traitements « inductifs » sont opaques non seulement pour les destinataires des décisions, mais également pour le responsable de traitement lui-même. Il arrive en effet que la machine, entraînée à partir d'un corpus qualifié ou poursuivant un objectif par renforcement, produise pour elle-même l'ensemble de règles qu'elle suppose le plus efficace, dans une forme qui n'est plus accessible à l'intelligence humaine[9]. Cette opacité *erga omnes* est alors sans remède.

Les dangers de ces « boîtes noires algorithmiques » ont été maintes fois dénoncés, à juste titre[10]. Mais une fois encore, le propos mérite d'être nuancé. Les processus décisionnels opaques ne sont pas apparus avec l'informatique. Des hommes étaient envoyés il y a quelques années encore en prison, et avant cela à l'échafaud, sur « l'intime conviction » d'un jury d'assises[11]. Cela a longtemps

8 Ce sont les techniques dont il est le plus souvent affirmé qu'elles relèvent de « l'intelligence artificielle ». V. toutefois, pour une définition bien plus englobante de l'IA, S. Merabet, Vers un droit de l'intelligence artificielle, préf. H. Barbier, Dalloz, Nouvelle bibliothèque de thèses n° 197, 2020. La présente étude ne fera pas appel au concept d'IA, qui ne fait actuellement consensus ni en droit ni en informatique. V. toutefois infra l'existence d'un projet de réglementation européenne en la matière.

9 V. not. J.-M. Deltorn, « Le droit des données personnelles face à l'opacité des algorithmes prédictifs : les limites du principe de transparence », in Regards sur le nouveau droit des données personnelles, éd. Ceprisca, 2019, p. 153.

10 Par ex. E. Mouriesse, « L'opacité des algorithmes et la transparence administrative », RFDA, 2019, p. 45 ; T. Douville, « Parcoursupr et le secret des algorithmes », Dalloz IP/IT, 2019, p. 700 ; G. Julia, « Intelligence artificielle et droit », Droit et Patr., 2020, n° 298 ; A. Penven, « La transparence des algorithmes et le consommateur », Cahiers de droit de l'entreprise, n° 5, sept. 2019, dossier 30 ; S. Merabet, thèse préc., n° 274 s. ; CNIL, Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle, rapport de 2017, p. 53.

11 L'exigence d'une motivation du principe de la condamnation par le jury d'assises a été introduite à l'article 365-1 du Code de procédure pénale par une loi n° 2011-939 du 10 août 2011. Une décision du Conseil constitutionnel

semblé acceptable parce qu'une procédure rigoureuse encadre la présentation et la discussion des preuves, et que la collégialité des débats mélange les subjectivités individuelles pour mieux les sublimer en une supposée objectivité collective. Il est possible de recruter ou de promouvoir un salarié, de choisir un locataire sans guère s'en expliquer, mais des règles probatoires favorables viendront *a posteriori* au secours de celui qui s'estimerait victime d'une discrimination[12].

**6. Présentation de l'encadrement actuel.** Au fond, ce que l'on attend d'un processus décisionnel, automatisé en tout, en partie ou pas du tout, c'est qu'il présente certaines *qualités* (d'équité, de justice) qu'on espère obtenir par la mise en place de *garanties*, variables selon les contextes (transparence de la procédure, motivation du choix, existence de recours…)[13].

Les insuffisances présentées par les décisions « purement humaines » ont pu être identifiées de longue date, et les garanties de qualité être ajustées en conséquence, dans chaque secteur considéré (celui du travail, des examens et concours, de la justice, du logement, de la banque…). En revanche, la montée en puissance relativement rapide des systèmes informatisés de décision ou d'aide à la décision a encouragé le législateur européen à se demander, de manière transverse, si ces systèmes présentent des dangers qui leur sont propres, et s'il convient en retour d'édicter des garanties qui leur soient spécifiques[14]. Le RGPD offrait pour cela un cadre cohérent, puisque les procédures informatiques visant à fixer les droits et obligations d'une personne donnée ont nécessairement pour carburant des données à caractère personnel. La richesse de l'identité individuelle est subitement ramenée à quelques indicateurs considérés comme pertinents par le responsable de traitement, afin qu'il en soit tiré une décision : c'est bien un traitement de données, soumis au Règlement[15].

De là vient l'article 22 du RGPD, qui dispose tout d'abord : « *1. La personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire* ». On entend par « profilage » le rattachement de la personne dont les données sont traitées à une catégorie, comme celle des « débiteurs fiables » ou des « salariés peu performants », ce qui peut être le préalable à une décision (respectivement d'octroi de crédit ou de refus de promotion, par exemple)[16].

Le texte tape fort, mais sur une cible relativement étroite. La cible est étroite, car seules sont concernées les décisions fondées « exclusivement » sur un traitement automatisé. Une lecture stricte de la disposition implique qu'à chaque fois que la décision se fonde sur au moins un autre élément

n° 2017-694 QPC du 2 mars 2018 a par la suite exigé que la motivation porte également sur le quantum de la peine.

12 V. not. l'art. L. 1132-3-3 C. trav., évoqué dans la suite de l'étude.

13 S'agissant des décisions informatisées, S. Desmoulin-Canselier et D. Le Métayer, op. cit., recherchent dans chaque domaine étudié les qualités suivantes : « légitimité », « efficacité, fiabilité », (absence de) « discrimination, biais », « contrôle, intelligibilité », « protection des données personnelles et de la vie privée ».

14 Sur l'ensemble de cette question, voir notamment, outre l'ouvrage précité de S. Desmoulins-Canselier et D. Le Métayer, spéc. p. 123 s. : J. Rochfeld, « Droit des personnes — Droit de ne pas subir une décision fondée exclusivement sur un traitement automatisé », in Droit des données personnelles. Les spécificités du droit français au regard du RGPD, dir. N. Martial-Braz et J. Rochfeld, p. 176 s. ; J. Rochfeld, « Données à caractère personnel — Droit de ne pas subir une décision fondée sur un traitement automatisé », in Rép. IP/IT et communication Dalloz, mai 2020 ; T. Douville, Droit des données à caractère personnel, Gualino, 2021, n° 507 s.

15 Les prises de décisions automatisées pourraient par ailleurs être encadrées par la proposition de règlement européen sur l'IA, à supposer qu'elle soit adoptée un jour (*Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence [artificial intelligence act] and amending certain union legislative acts, 21/04/21)*. L'IA y est en effet définie à l'article 3.1 comme « s*oftware that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with* ».

16 Sur cette notion : N. Martial-Braz, "Le profilage. Fiche pratique », CCE, 2018, n 4, dossier 15.

que le traitement automatisé, elle sort entièrement du champ d'application du texte, quand bien même l'apport de l'automatisme serait prépondérant, voire écrasant par rapport aux autres sources mobilisées. Mais lorsque le processus décisionnel repose tout entier sur un traitement algorithmique, le texte tape fort : même si des exceptions sont prévues le principe est l'interdiction.

En droit français, c'est l'article 47 de la loi informatique et libertés qui dispose :

« *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.*

*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 (...)* ».

L'approche française de la question présente ainsi plusieurs spécificités. D'abord, un premier alinéa qui pose une interdiction spécifique en matière de décisions de justice, à rebours des positions estoniennes. Dès lors qu'un traitement est utilisé pour « évaluer certains aspects de la personnalité », il ne peut servir à fonder la décision, même partiellement[17]. Ensuite, au deuxième alinéa, nous laissons derrière nous le domaine spécifique des décisions de justice pour rencontrer une transcription de la solution générale de l'article 22 du RGPD. Mais elle n'est pas exprimée, comme dans le règlement, sous la forme d'un droit subjectif de la personne concernée « à ne pas faire l'objet » de certains types de décisions[18].

Revenons au texte européen. L'interdiction de principe connaît trois exceptions[19]. La décision reposant exclusivement sur un traitement automatisé peut être autorisée par le droit de l'Union ou le droit national du responsable de traitement, moyennant des « mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée » (b). Cette marge de manœuvre a notamment été utilisée par la France en matière de décisions administratives[20]. La décision est également permise si elle est « nécessaire à la conclusion ou à l'exécution d'un contrat entre la personne concernée et un responsable du traitement » (a) ou qu'elle est fondée sur le « consentement explicite de la personne concernée » (c). Alors que l'exception (b) déléguait entièrement aux droits européens et nationaux le soin d'édicter des garanties adéquates, les cas (a) et (c) déclenchent l'application d'une protection-socle[21]. En effet, « (…) le responsable du traitement met en œuvre des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée, **au moins** du droit de la personne concernée d'obtenir

17 L. Huttner, « Données à caractère personnel – Décision automatisée et justice », Répertoire Dalloz IP/IT et Communication, n° 14, relève que cette position française aboutit à un encadrement plus sévère des décisions automatisées que celui du RGPD, sur une question pour laquelle le règlement ne semble pas ouvrir de marge de manœuvre nationale. La question de la conformité au droit européen est donc posée.

18 Sur la différence entre la consécration d'un droit subjectif et l'édiction d'une interdiction : J. Rochfeld, art. préc. in Rép. Dalloz IP/IT, n° 13 . T. Douville, op. cit., n° 573.

19 Art. 22, 2° du RGPD.

20 L'article 47 de la loi informatique et libertés prévoit ainsi le cas « (…) 2° Des décisions administratives individuelles prises dans le respect de l'article L. 311-3-1 et du chapitre Ier du titre Ier du livre IV du code des relations entre le public et l'administration, à condition que le traitement ne porte pas sur des données mentionnées au I de l'article 6 de la présente loi. Ces décisions comportent, à peine de nullité, la mention explicite prévue à l'article L. 311-3-1 du code des relations entre le public et l'administration. Pour ces décisions, le responsable de traitement s'assure de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard ». La suite de l'étude se concentrant sur l'amélioration des textes au niveau européen, cette réglementation française spécifique ne sera pas développée.

21 Art. 22, 3° du RGPD.

une intervention humaine de la part du responsable du traitement, d'exprimer son point de vue et de contester la décision ».

L'article 22 du RGPD s'achève par des prévisions spécifiques aux traitements « qui révèle(nt) 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 »[22]. Pour qu'ils constituent le fondement exclusif d'une décision produisent des effets juridiques, ou affectant significativement la personne de manière similaire, des conditions supplémentaires sont exigées[23].

Les décisions automatisées font par ailleurs l'objet d'un encadrement par les articles 13 et 14 du règlement. Ces textes imposent de communiquer aux personnes concernées un certain nombre d'informations relatives à l'identité du responsable de traitement, à la finalité poursuivie, aux durées de conservation, etc. Mais ils commandent également de signaler « l'existence d'une prise de décision automatisée, y compris un profilage, visée à l'article 22, paragraphes 1 et 4, et, au moins en pareils cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée »[24].

**7. Critique et propositions.** Cet encadrement européen des décisions automatisées est fort bienvenu dans son principe. Mais est-il satisfaisant dans ses modalités ? Nous essaierons de démontrer que la rédaction actuelle de l'article 22 est critiquable sur plusieurs points (I) et formulerons des propositions d'amélioration (II).

# I — Une critique de l'article 22 du RGPD

Le texte de l'article 22 apparaît perfectible tant du point de vue des décisions visées (A) que des garanties déployées lorsqu'il est admis, par exception, qu'une décision soit prise sur le fondement exclusif d'un traitement automatisé (B).

## A – Des critères inadéquats

Le règlement vise les « décisions » produisant des effets juridiques ou affectant la personne concernée de façon similaire prises « sur le fondement exclusif » d'un « traitement automatisé ». La notion de « décision » appelle de simples précisions (1). En revanche, le critère du « traitement automatisé » est trop large, frappant d'une interdiction de principe des traitements qui ne méritent

22 Art. 22, 4° du RGPD.

23 Il faudra non seulement se trouver dans le champ de l'une des trois exceptions précédemment citées (autorisation par le droit de l'Union ou un droit national, nécessité pour l'exécution d'un contrat ou consentement explicite de la personne), mais également se prévaloir de l'article 9, a) ou g). L'article 9, a) dispose : « la personne concernée a donné son consentement explicite au traitement de ces données à caractère personnel pour une ou plusieurs finalités spécifiques, sauf lorsque le droit de l'Union ou le droit de l'État membre prévoit que l'interdiction visée au paragraphe 1 ne peut pas être levée par la personne concernée ». L'article 9, g) dispose : « le traitement est nécessaire pour des motifs d'intérêt public important, sur la base du droit de l'Union ou du droit d'un "État membre qui doit être proportionné à l'objectif poursuivi, respecter l'essence du droit à la protection des données et prévoir des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée ».

24 Articles 13, 2, f) et 14, 2, g) du RGPD.

pas une telle défiance (2). Le critère du « fondement exclusif » est quant à lui trop étroit, ce qui permet à des automatismes potentiellement dangereux d'échapper au texte (3).

## 1 — La « décision » : une notion à préciser

**8. Présentation**. Certains éléments du champ d'application de l'article 22 paraissent devoir être repensés. Ce n'est peut-être pas le cas de la notion de « décision », qui appelle seulement quelques remarques interprétatives. Elle n'opère pas de manière autonome : le règlement vise les décisions « produisant des effets juridiques » à l'égard de la personne concernée ou « l'affectant de manière significative de façon similaire ».

**9. Interprétation extensive : le caractère sécable des décisions**. Observons tout d'abord qu'il peut être nécessaire, afin de conserver au texte une efficacité maximale, de scinder une prise de position apparemment globale du responsable de traitement en plusieurs petites « décisions ». Imaginons une banque qui ferait appel à un arbitrage humain afin de déterminer à qui elle accorde ou refuse des crédits, mais qui laisserait ensuite un automatisme fixer le taux d'intérêt pour les opérations acceptées, sans supervision. S'en tenir à une analyse globale laisserait penser qu'« une décision » de prêt à 1,25 % a été prise de concert par l'Homme et la machine. Il faut cependant restituer au montage son exacte qualification : ce sont deux décisions qui ont été prises, l'une sur le principe et l'autre sur le taux. La seconde a donc été entièrement automatisée : nous verrons plus loin quel régime juridique appelle ce constat.

**10. Interprétation restrictive : une procédure visant à fixer une situation individuelle.** Il faut donc veiller à trouver dans une situation donnée autant de décisions qu'il est nécessaire. Cela ne signifie pas qu'il faut voir des décisions partout. Après avoir donné de l'élan au concept, cherchons à le cantonner à de justes proportions. Imaginons un commerçant dont l'entreprise apparaîtrait en position très favorable dans les résultats de Google Search à la suite d'une requête très populaire (comme les mots « automobile d'occasion » suivis du nom d'une ville). Voici que la firme de Mountain View révise son traitement algorithmique, comme elle le fait régulièrement : le commerçant est tout à coup relégué à la deuxième page des résultats, sa visibilité s'effondre et son chiffre d'affaires avec elle. Peut-on considérer que l'agencement des hyperliens en réponse à une recherche constitue une décision ? On pourrait chercher à contourner la difficulté en s'appuyant sur le fait qu'il ne s'agit pas, en toute hypothèse, d'une décision « produisant des effets juridiques ». Dans le cadre des résultats « naturels », par opposition aux résultats « sponsorisés », le moteur de recherche n'est pas en relation contractuelle avec la personne concernée, et l'hypothèse étudiée n'engage pas sa responsabilité civile extracontractuelle. Mais qu'en est-il des décisions « affectant (la personne) de manière significative de façon similaire » ? N'est-ce pas le cas ici ? Le considérant 71 du RGPD cite en exemples « le rejet automatique d'une demande de crédit en ligne ou des pratiques de recrutement en ligne sans aucune intervention humaine ». Refusant à des personnes le bénéfice d'un contrat dont elles recherchent la conclusion, ces décisions pourraient pratiquement être considérées comme produisant des effets juridiques. Les décisions affectant la personne « de façon similaire » seraient alors des décisions quasi juridiques ou para-juridiques. Mais le CEPD en propose une interprétation plus large, considérant que « le niveau d'importance doit être similaire à celui d'une décision produisant un effet juridique »[25]. Une telle décision, ajoute-

25 CEPD, Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679, 2018, p. 24.

t-il, « doit être de nature à affecter de manière significative la situation, le comportement ou les choix des personnes » et « avoir un impact prolongé ou permanent sur la personne concernée ». Tel pourrait être le cas, nous semble-t-il, de la relégation d'une entreprise en deuxième page des résultats de recherche de Google[26]. Le champ d'application de l'article 22 ainsi conçu apparaîtrait cependant trop large. Mais il est possible de resserrer la définition de la « décision » pour l'éviter. Postulons qu'une décision *a pour objet* de maintenir ou d'altérer la situation de la personne concernée. Les traitements algorithmiques produisant de *simples effets incidents* sur les individus comme dans notre exemple sont alors exclus. En ce qui les concerne, d'autres corps de règles peuvent intervenir pour le cas où l'automatisme causerait un dommage illicite — notamment la responsabilité civile extracontractuelle — mais l'article 22 du RGPD ne constitue pas un outil adéquat. À titre d'exemple, en revanche, l'utilisation d'outils de modération automatisée déclenchant le retrait d'un contenu sur les réseaux sociaux, voire une simple dégradation de sa visibilité, constituerait bien une « décision » : le processus a pour objet de produire un verdict individuel[27].

La notion de « décision » appelait donc quelques précisions, mais ne suscite par elle-même aucune difficulté insurmontable. Il en va autrement de l'expression « traitement automatisé ».

## 2 — Les « traitements automatisés » : une notion trop large

**11. La méfiance de principe envers tout processus entièrement automatisé.** Supposons qu'il faille prendre une décision produisant des effets juridiques ou affectant l'individu de manière significative de façon similaire. Surgit alors une interdiction — fût-elle formulée comme un droit subjectif de la personne concernée — de l'adosser entièrement à un « traitement automatisé ». Il existe certes des exceptions, à la remorque desquelles se trouve un cortège de garanties spécifiques. Mais le principe est bien la prohibition. Il en découle deux corollaires. Premièrement : laisser le dernier mot à l'automatisme est un mal en soi, qui ne doit être toléré que dans de rares cas et moyennant d'importantes précautions. Deuxièmement : à l'inverse, placer un agent humain en bout de chaîne décisionnelle est nécessairement vertueux, puisqu'alors l'article 22 n'a plus rien à dire.

Parce qu'elles sont trop générales, ces affirmations sont erronées et susceptibles, en certains cas, de produire des effets pervers.

Elles s'expliquent fort bien si l'on suppose que le législateur européen, en édictant ces règles, visait essentiellement les traitements algorithmiques inductifs, ceux qui sont susceptibles de sécréter leur propre norme, par exemple par apprentissage machine à partir de données d'apprentissage qualifiées. Comme l'affirmait Jean Foyer dès 1977, en tant que rapporteur de la loi informatique et libertés pour l'Assemblée nationale, « l'ordinateur n'a, dit-on, aucune "faculté d'étonnement" devant les erreurs de droit ou de fait, qui peuvent affecter les données. Il n'en a point davantage

26 On relèvera d'ailleurs qu'en droit français, l'art. 47 de la loi informatique et libertés vise une décision « produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative », sans reprendre la formule « de façon similaire ». Comp. T. Douville, op. cit., n° 581 : « La décision doit affecter la personne concernée, c'est-à-dire produire des effets similaires à des effets juridiques ».

27 Il s'agit par ailleurs d'une décision produisant des effets juridiques, à un double titre : en ce qu'elle est la sanction de la norme violée par l'utilisateur (conditions générales, ou règle étatique) et touche à une liberté fondamentale (la liberté d'expression). Il n'est donc pas surprenant que le projet de règlement relatif à un marché intérieur des services numériques du 15 décembre 2020 prévoie que l'utilisation d'outils de modération automatisés doive être signalée à l'utilisateur (art. 14,6, 15,2 [c]) et, surtout qu'il proscrive qu'une décision en la matière « soi [t] uniquement prise par des moyens automatisés » (art. 17,5).

devant les déductions fausses imputables au programme »[28]. Si un automatisme doit apprendre à distinguer les chats des chiens, mais qu'il se trouve dans son corpus d'apprentissage un nombre écrasant de chats à pelage clair et de chiens à pelage sombre, il pourrait fort bien conclure qu'un animal de couleur claire est un chat. Plus inquiétant : si un traitement chargé de repérer des CV prometteurs apprend à partir d'un corpus dans lequel les trajectoires professionnelles « réussies » sont essentiellement celles d'hommes blancs, parce que le racisme et la misogynie ont artificiellement promu ces profils au détriment des autres au cours des décennies écoulées, l'ordinateur intégrera ces données — c'est ce qui est arrivé à Amazon[29]. Il est donc essentiel de se méfier de tels algorithmes inductifs et d'éviter de leur laisser le dernier mot : dans ces cas de figure, l'intervention d'une « faculté d'étonnement » humaine en dernière analyse est nécessaire, et le dispositif de l'article 22 absolument justifié.

**12. L'innocuité de certains processus entièrement automatisés**. Mais considérons à présent un traitement algorithmique purement déductif, qui se contenterait de mettre en œuvre mécaniquement des critères délibérés par avance selon des procédés décisionnels humains : la machine ne fait naître aucun danger qui lui soit propre. Attaquer l'informatique ici, c'est tirer sur le messager.

Prenons l'exemple des algorithmes publiés par la Ville de Nantes, et qui servent pour l'un à calculer la tarification sociale des transports, pour l'autre la tarification sociale de l'eau[30]. Les règles qui les régissent ont été adoptées par les organes politiques compétents, dans les formes requises. La teneur en est publiée, de sorte que chacun puisse vérifier que la traduction de la langue du droit dans la langue des machines est absolument fidèle[31]. Pour chaque administré, le traitement est en mesure de produire une décision sans qu'aucune intervention humaine ne soit plus nécessaire. Il s'agit bien d'une décision au sens où nous l'entendions, à savoir une procédure ayant pour objet d'effectuer un choix destiné à maintenir ou à modifier la situation d'un individu. Cette décision produit des effets juridiques, puisqu'elle module le montant des créances dont est titulaire la ville à l'égard du consommateur d'eau.

Observons par ailleurs qu'un fonctionnaire humain pourrait appliquer la formule lui-même, muni d'un papier et d'un crayon : le gâchis de ressources serait phénoménal sans que la qualité des résultats soit susceptible de progresser en rien[32]. Il en ira de même pour la quasi-totalité des systèmes de facturation publics et privés, qui appliqueront un barème réglementaire ou contractuel par un moyen purement automatisé afin d'en tirer une décision individuelle. La notation d'un examen réalisé sous forme d'un QCM, pour lequel un barème officiel a été adopté, obéit à une logique similaire.

---

28 Compte-rendu intégral de la séance de l'Assemblée nationale du mardi 4 octobre 1977, JO, p. 5782.

29 J. Dastin, « Amazon scraps secret AI recruiting tool that showed bias against women », article reuters.com du 11 octobre 2018 : « That is because Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry. In effect, Amazon's system taught itself that male candidates were preferable. It penalized resumes that included the word 'women's,' as in 'women's chess club captain'. »

30 https://data.nantesmetropole.fr/pages/algorithmes_nantes_metropole/.

31 Sur la transcription d'un algorithme décrit par un texte juridique en code informatique : Liane Huttner, Denis Merigoux. « Traduire la loi en code grâce au langage de programmation. Intelligence artificielle et finances publiques », Oct 2020, Nice, France, hal-03128248.

32 Relevons d'ailleurs qu'un tel traitement manuel peut encore être considéré comme « automatisé », selon le sens que l'on retient pour ce terme. Dans le Larousse en ligne, la première définition pour « automatisme » est certes « Qualité des appareils et installations qui fonctionnent sans intervention humaine », mais la deuxième est « Acte accompli mécaniquement, sans intervention de la volonté ».

Dans ces situations, les dispositions de l'article 22 apparaissent inadaptées. Le texte encourage le décideur à introduire une subjectivité humaine après le passage de l'automatisme, afin d'échapper à son champ d'application. Mais si la règle commune est juste, il est non seulement inopportun, mais même dangereux qu'un agent humain puisse y déroger pour un cas individuel, et cela sera quoi qu'il en soit juridiquement impossible dans la plupart des situations visées. Si la règle commune est problématique, elle doit être remise en cause par les voies de droit adaptées, mais la difficulté ne provient pas de la présence d'un traitement algorithmique, qui n'est ici qu'un fidèle exécutant, un avatar plus sophistiqué du boulier ou de la calculatrice.

La notion de « traitement automatisé » apparaît donc excessivement large. Au contraire, celle de « fondement exclusif » de la décision est trop étroite.

## 3 — Les décisions prises « sur le fondement exclusif » d'un traitement automatisé : un critère trop étroit

**13. Les systèmes d'aide à la décision**. Portons à nouveau notre attention sur des automatismes de traitement de l'information se fondant sur des critères imprévisibles et potentiellement inconnaissables, par exemple ceux qui exploitent certaines techniques d'apprentissage machine. De tels outils peuvent se révéler extrêmement précieux. À titre d'exemple, des logiciels d'analyse d'imagerie médicale peuvent détecter des anomalies si subtiles qu'elles échappent à l'œil humain. Il s'agira ici d'attirer l'attention du médecin, qui reprendra alors le contrôle du diagnostic. C'est lui, et lui seul, qui prendra la décision de retenir une thérapeutique. Aurait-on l'idée folle de laisser l'automatisme opter lui-même pour un traitement médicamenteux ou chirurgical, on se heurterait à l'interdiction de principe de fonder cette décision — qui ne produit pas des effets juridiques, mais affecte la personne de manière significative de façon similaire — sur un traitement automatisé. Autrement dit, si l'article 22 du RGPD encadre les décisions entièrement automatisées, il ne fait pas obstacle aux simples systèmes « d'aide à la décision » (SAD).

**14. Le danger : des décisions automatisées déguisées en SAD**. Mais l'on voit aussitôt poindre le danger. Prenez un automatisme qui, l'instant précédent, produisait à la chaîne d'authentiques décisions. Affichez à présent l'information qu'il produit sur un écran, devant lequel vous placerez un subordonné qui y jettera un œil distrait avant d'appuyer systématiquement sur un bouton vert : êtes-vous face à un simple outil de « monitoring », de « prétri » échappant à l'article 22[33] ?

Le CEPD l'a bien compris, qui retient du texte une interprétation plus exigeante. C'est seulement lorsqu'« un être humain examine et tient compte d'autres facteurs dans la prise de décision finale » qu'on est face à un simple SAD[34]. « Pour qu'il y ait intervention humaine, le responsable du traitement doit s'assurer que tout contrôle de la décision est significatif et ne constitue pas qu'un simple geste symbolique. Le contrôle devrait être effectué par une personne qui a l'autorité et la compétence pour modifier la décision. Dans le cadre de l'analyse, il convient de tenir compte de toutes les données pertinentes ».

On se heurte toutefois à une faiblesse inhérente à la logique globale de « compliance » dans laquelle le droit des données personnelles a basculé depuis l'entrée en application du RGPD : la supercherie risque de n'être révélée qu'au moment d'une des (rares) inspections de la CNIL. Certes, le Comité

33 Sur cette question, V. not. J. Rochfeld, art. préc. in Rép. Dalloz IP/IT, n° 7 s.
34 Lignes directrices précitées, p. 23.

estime que « Dans le cadre de son analyse d'impact relative à la protection des données, le responsable du traitement devrait identifier et consigner le degré d'intervention humaine dans le processus de prise de décision et le stade auquel cela se produit ». Mais le contenu de cette analyse d'impact, qui n'est pas publiée, restera soustrait aux regards des personnes concernées[35].

**15. Obligation de justifier l'apport propre du décideur humain**. Il semble envisageable de procéder autrement. Une simple aide à la décision, si elle est en principe exclue du champ de l'article 22, n'en constitue pas moins un traitement de données personnelles soumis au reste du RGPD. Or, l'introduction de cette étude a rappelé l'existence des articles 13 et 14, relatifs aux obligations de transparence. Lorsque le responsable de traitement prend des décisions automatisées, il doit en informer les personnes concernées et les éclairer sur « la logique sous-jacente » du traitement ainsi que sur ses conséquences. Ces textes pourraient être modifiés pour faire naître une obligation d'information spécifique aux traitements dont il est allégué qu'ils constituent de simples outils d'aide à la décision[36]. Cette information porterait sur l'apport propre du décideur humain, en particulier sur les critères et méthodes qu'il emploie.

À défaut d'une telle précision, le prétendu système d'aide à la décision serait présumé fonctionner exclusivement sur le fondement d'un traitement automatisé.

À supposer que le responsable de traitement défère à cette obligation d'information, mais de manière particulièrement elliptique, il y aurait là matière à alerter les personnes concernées ou l'autorité de contrôle, là où elles peuvent actuellement être tenues dans une ignorance complète.

Quant au responsable de traitement qui aurait indiqué un apport humain spécifique qui ne correspondrait pas à la réalité de ses pratiques, il serait passible des lourdes sanctions administratives édictées par le RGPD en cas de non-respect des droits des personnes concernées.

Après avoir proposé plusieurs améliorations s'agissant des critères délimitant le champ d'application de l'article 22, examinons à présent les garanties dont bénéficient les personnes concernées lorsque, par exception, une décision purement automatisée est admise.

## B – Des garanties imprécises

**16. Le triptyque des garanties**. L'interdiction de prendre des décisions lourdes d'enjeux exclusivement sur le fondement d'un traitement automatisé n'est qu'un principe, qui souffre des exceptions. Laissons de côté celles qui seraient prévues par le droit de l'Union ou de l'État membre moyennant des garanties dont on sait seulement qu'elles devront être « appropriées ». S'agissant des deux autres exceptions. la nécessité pour la conclusion ou l'exécution d'un contrat et le consentement explicite de la personne concernée., un socle de garanties est en revanche précisé par le RGPD : il s'agit « au moins du droit de la personne concernée d'obtenir une intervention humaine de la part du responsable du traitement, d'exprimer son point de vue et de contester la décision ».

35 En vertu de l'article 36 du RGPD, même l'autorité de contrôle n'aura connaissance de l'AIPD que si le responsable de traitement estime, au terme de son analyse, que les risques pour les personnes sont élevés. Le risque qu'il soit clément avec lui-même n'est pas nul.

36 En ce sens : T. Douville, op. cit., n° 581.

La formule manque de clarté. Elle est articulée autour de trois éléments. Leur ordre est-il significatif ? S'agit-il des étapes d'un processus global, classées par ordre chronologique ? Mais alors, en quoi consiste « l'intervention humaine » qui prend place avant même que la personne ait exprimé son point de vue ? Le fait d'exprimer son point de vue vient-il nécessairement à l'appui d'une contestation de la décision ? Dans le cas contraire, à quoi sert-il d'exprimer un point de vue critique, si la révision n'est pas sollicitée ?

Le considérant 71 du règlement est à peine plus éclairant : « En tout état de cause, un traitement de ce type devrait être assorti de garanties appropriées, qui devraient comprendre une information spécifique de la personne concernée ainsi que le droit d'obtenir une intervention humaine, d'exprimer son point de vue, d'obtenir une explication quant à la décision prise à l'issue de ce type d'évaluation et de contester la décision ». La notion « d'information spécifique » n'est pas explicitée, et les lignes directrices du CEPD ne se hasardent pas à la définir. En revanche, l'apport de ce considérant est d'affirmer que la personne concernée a droit à une *explication* de la décision prise à son endroit.

**17. Les composantes de « l'intervention humaine »**. Tout semble finalement plus clair si l'on considère « l'intervention humaine » non pas comme l'une des garanties figurant parmi d'autres dans la liste, mais comme la clé de voûte de l'ensemble. Soit une personne à qui l'on vient de notifier une décision rendue par une machine agissant seule. L'individu considère avoir été traité injustement. La première chose à faire, c'est de placer un interlocuteur humain face à lui, car on ne négocie pas avec un ordinateur.. pour reprendre une formule de l'ingénieur Benjamin Bayart, « l'ordinateur est fatal »[37].

Une large palette d'interactions est alors possible entre les deux personnes ainsi mises en relation. Dans un ordre chronologique cohérent, on pourrait d'abord envisager que l'agent humain du responsable de traitement réponde à des questions, explique la décision, fasse œuvre de pédagogie. Encore faut-il que ce soit possible ! Tout l'intérêt de certaines décisions automatisées, on l'a dit, c'est d'être produites en relative autonomie par la machine, selon une logique dont les détails nous restent inaccessibles. Ensuite, l'explication — ou l'absence d'explication — ayant éventuellement échoué à convaincre le destinataire de la décision, celui-ci peut en demander la révision. Il reviendra alors à l'humain de rouvrir l'instruction, en quelque sorte. Relevons au passage que si un premier agent a été chargé d'expliquer — et donc, indirectement, de défendre — la décision automatisée, il n'occupe sans doute plus une position suffisamment impartiale pour statuer en cause d'appel. L'article 22 attend de cette seconde décision qu'elle soit de surcroît contradictoire, c'est-à-dire que la personne concernée devra avoir eu l'opportunité de présenter, si elle le souhaite, des éléments au soutien de sa cause.

Des pistes d'amélioration ont ainsi été proposées tant du point de vue du champ d'application de l'article 22 que des garanties offertes aux personnes concernées lorsqu'une décision entièrement automatisée est rendue. Voyons à présent comment ces suggestions peuvent être exploitées pour réorganiser, voire réécrire en partie le texte.

37 V. par ex. dans B. Bayart et M. Rees, « Pass sanitaire, géopolitique de la Data, copie privée ? », conférence Thinkerview du 15 juin 2021, https://www.youtube.com/watch?v=EOWeewlc2CE.

# II — Une refonte de l’article 22 du RGPD

Il est proposé d’instaurer deux régimes juridiques profondément distincts selon que le traitement automatisé à l’origine de la décision est « transparent » (A) ou « opaque » (B). Le premier constitue un régime de faveur, le second un régime de prudence.

## A – Le régime de faveur des traitements décisionnels transparents

**18. Une autorisation de principe.** Dans la première partie de l’étude, nous nous sommes appuyés sur l’exemple de la tarification sociale de l’eau de la ville de Nantes pour montrer l’absence de danger que présente un traitement automatisé de pure exécution, qui ne fait qu’appliquer une décision-cadre à l’échelle individuelle.

L’approche actuelle déclenche l’interdiction de principe et contraint à se réfugier derrière l’une des exceptions prévues par l’article 22, 2°. Il est certainement possible de considérer que de tels automatismes sont fondés sur le droit national lorsqu’ils relèvent du secteur public, ou sur la nécessité pour l’exécution du contrat s’agissant du secteur privé. Mais il ne devrait pas être nécessaire à une société commerciale qui souhaite appliquer mécaniquement ses tarifs publics à un client donné de justifier que cela est *strictement nécessaire* à l’exécution d’une convention. L’exiger reviendrait à affadir cette exception, alors qu’elle devra être disponible dans toute sa puissance s’agissant des traitements opaques dont il sera question plus bas. Il ne doit tout simplement pas y avoir d’interdiction de principe d’appliquer de tels automatismes sans intervention humaine systématique.

**19. Définition des traitements transparents**. Encore faut-il délimiter correctement le champ de ces traitements qu’on prétend inoffensifs, ou du moins exempts des dangers qui caractérisent les traitements opaques. Ils pourraient être définis ainsi :

« Un traitement automatisé destiné à produire une décision individuelle est dit “transparent” lorsque la nature et la pondération des variables en entrée sont intégralement fixées et décrites, avant son exécution, dans une décision-cadre dont le contenu est accessible aux personnes concernées, accompagnée au besoin d’une explication rédigée dans un langage simple et clair, et dont les décisions individuelles constituent la simple exécution ».

**20. Droit à une explication et traitements transparents**. Admettons que les automatismes ainsi définis doivent échapper à l’interdiction de principe qui frappera les traitements opaques. Dans le schéma actuel, c’est seulement lorsque le traitement était en principe interdit, mais que l’on se trouve dans le champ d’une exception que la personne concernée bénéficie d’un socle de garanties. Mais rien n’interdit de changer complètement d’approche, et d’édicter directement un droit à une « intervention humaine » s’agissant des traitements transparents. Simplement, la nature de cette intervention doit être bien comprise. Par hypothèse, il n’y aurait aucun sens ici à former un recours : en rendant la seconde décision, l’être humain resterait sous la dépendance du barème, du tarif ou du cadre dont la décision première n’était qu’une application circonstancielle. Pour les mêmes raisons, proposer à la personne concernée de présenter des observations à l’appui de son appel afin d’assurer un caractère contradictoire à la procédure serait totalement vain. On pourrait certes imaginer du responsable de traitement qu’il reprenne la formule avec la personne concernée pour lui en

expliquer le sens, mais si cette formule est complexe, il devrait avoir déjà publié ou avoir rendu accessible une telle explication au bénéfice de tous. L'intérêt d'une intervention humaine individuelle est alors de vérifier et de commenter l'application du barème ou du cadre aux valeurs spécifiques qui ressortent du dossier individuel de la personne concernée. À titre d'exemple, lorsque les services fiscaux sont interrogés sur le montant d'une taxe d'habitation, ils rappellent la formule de calcul et, pour chacun de ses éléments, donnent la valeur qui a été retenue au cas d'espèce, ce qui permet à l'intéressé de constater que le total ne comporte aucune erreur[38].

« Le droit à une intervention humaine », dans le cadre d'un traitement décisionnel « transparent », pourrait donc être formulé comme suit :

« Il est communiqué à la personne concernée, à sa demande, l'état des variables employées pour produire la décision individuelle qui lui est opposée, de manière à ce qu'elle puisse vérifier que la décision-cadre lui a été appliquée sans erreur ».

**21. Remise en cause de la décision-cadre**. Si c'est la teneur de la décision-cadre elle-même qui paraît problématique aux destinataires de décision, la réponse ne relève pas du RGPD. Selon le domaine considéré, il convient d'invoquer le droit administratif, civil, commercial, du travail ou de la consommation pour obtenir que le cadre prétendument injuste soit amendé. Le fait qu'un tarif contractuel ou réglementaire fasse l'objet d'une application individuelle automatisée ne change rien à sa nature.

**22. Contrôle des discriminations**. Qu'en est-il, enfin, du risque de décision discriminatoire[39] ? Par hypothèse, il sera facile de vérifier *ex ante* qu'un tel traitement n'opère aucune discrimination directe, puisque les critères employés sont intégralement publics. Il reste cependant envisageable qu'il produise des effets de discrimination indirects[40]. Il est alors possible de s'inspirer de l'article L. 1134-1 du Code du travail français, en procédant à des vérifications *ex post*, en aménageant la charge de la preuve : si la personne concernée présente des éléments de fait laissant supposer que le traitement automatisé transparent est à l'origine de discriminations indirectes fondées sur un critère prohibé, il incombe au responsable de prouver que sa décision n'est pas fondée sur de tels critères[41].

38 Expérience menée par l'auteur.

39 Sur ce sujet : Défenseur des droits (en partenariat avec la CNIL), Algorithmes : prévenir l'automatisation des discriminations, rapport de 2020 ; Cédric Villani (dir.), Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne, p. 147, qui propose l'instauration d'un *Discrimination Impact Assessment* : « . Il s'agirait d'accompagner le PIA d'un dispositif analogue pour les discriminations, un d*iscrimination impact assessment*, ou DIA, pour amener les concepteurs d'IA à s'interroger sur les conséquences sociales des algorithmes qu'ils produisent » ; G. Haas et S. Astier, « Les biais de l'intelligence artificielle : quels enjeux juridiques ? », Répertoire Dalloz IP/IT et Communication, juillet 2019.

40 Au sens de l'article 1er de 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 : « Constitue une discrimination indirecte une disposition, un critère ou une pratique neutre en apparence, mais susceptible d'entraîner, pour l'un des motifs mentionnés au premier alinéa, un désavantage particulier pour des personnes par rapport à d'autres personnes, à moins que cette disposition, ce critère ou cette pratique ne soit objectivement justifié par un but légitime et que les moyens pour réaliser ce but ne soient nécessaires et appropriés ».

41 Art. L. 1134-1 al. 1 C. trav. : « Lorsque survient un litige en raison d'une méconnaissance des dispositions du chapitre II, le candidat à un emploi, à un stage ou à une période de formation en entreprise ou le salarié présente des éléments de fait laissant supposer l'existence d'une discrimination directe ou indirecte, telle que définie à l'article 1er de 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 critères ne devant faire l'objet d'aucune discrimination devrait soit être abandonnée aux États, soit refondue au niveau européen. Actuellement, le RGPD se concentre sur les catégories particulières de données de l'article 9, ce qui ne permet pas de prendre en compte, par exemple, la discrimination pourtant classique fondée sur la simple appartenance au genre féminin ou masculin[42].

Dans le cadre des traitements automatisés opaques, en revanche, des discriminations directes sont tout à fait possibles. De manière générale, leur dangerosité est supérieure, ce qui justifie qu'il leur soit adossé un régime juridique beaucoup plus contraignant.

## B – Le régime strict des traitements décisionnels opaques

**23. Définition des traitements opaques**. Que faut-il entendre par « traitements opaques » ? Leur définition peut être conçue très simplement en renversant celle des traitements transparents. Alors que les traitements transparents laissent apparaître la nature et la pondération de toutes les variables employées, les traitements sont opaques quand tout ou partie de ces informations reste inaccessible aux destinataires des décisions. Quand cela se produit-il ?

D'abord, lorsque certaines familles d'automatismes inductifs sont employées. Nous avons vu qu'il existe des techniques d'apprentissage dans lesquelles les critères exacts finalement utilisés pour le classement sont inaccessibles à l'utilisateur, et même au concepteur.

À cette opacité intrinsèque et irrémédiable s'ajoutent, ensuite, des situations dans lesquelles les critères de décision sont connus, mais volontairement passés sous silence. Quand bien même on utiliserait un automatisme déductif, qui se contenterait de déployer un ensemble de règles conçues par l'esprit humain, ou des traitements inductifs capables de rendre compte de façon intelligible des choix qu'ils opèrent, le responsable de traitement ou le concepteur du logiciel fait parfois le *choix* de l'obscurité. Il existe des raisons légitimes de le faire, par exemple afin de protéger les détails d'un processus décisionnel original, dont l'efficacité procure un avantage concurrentiel, et dans lequel ont été investis du temps et des moyens. Mais l'obscurité peut également être un moyen de dissimuler l'inavouable, comme l'utilisation d'un critère discriminatoire prohibé. Dans tous les cas, l'opacité doit être assumée, et produira des conséquences juridiques.

Cette approche a de surcroît l'avantage d'être technologiquement neutre, comme l'est le RGPD de façon générale. Ce serait une erreur de s'arrimer à un état transitoire de la technique, en visant par exemple des familles précises de *machine learning* ou « d'intelligence artificielle », dont les contours seraient à la fois contestables et évolutifs.

La définition pourrait donc être libellée ainsi : « Un traitement automatisé destiné à produire une décision individuelle est dit "opaque" lorsque le responsable de traitement prive la personne

42 Cela ressort très bien de la lecture du considérant 71 : « Afin d'assurer un traitement équitable et transparent à l'égard de la personne concernée, compte tenu des circonstances particulières et du contexte dans lesquels les données à caractère personnel sont traitées, le responsable du traitement devrait utiliser des procédures mathématiques ou statistiques adéquates aux fins du profilage, appliquer les mesures techniques et organisationnelles appropriées pour faire en sorte, en particulier, que les facteurs qui entraînent des erreurs dans les données à caractère personnel soient corrigés et que le risque d'erreur soit réduit au minimum, et sécuriser les données à caractère personnel d'une manière qui tienne compte des risques susceptibles de peser sur les intérêts et les droits de la personne concernée et qui prévienne, entre autres, les effets discriminatoires à l'égard des personnes physiques **fondées sur la l'origine raciale ou ethnique, les opinions politiques, la religion ou les convictions, l'appartenance syndicale, le statut génétique ou l'état de santé, ou l'orientation sexuelle,** ou qui se traduisent par des mesures produisant un tel effet » (souligné par nous).

concernée d'une description exhaustive des règles qui lui sont appliquées, que ce soit par choix ou en conséquence des techniques employées par le traitement ».

**24. Maintien de l'interdiction de principe**. Les traitements opaques ayant été définis, on peut leur appliquer pleinement l'interdiction de principe d'être exclusivement à l'origine des décisions les plus lourdes d'enjeux — celles produisant des effets juridiques ou affectant significativement la personne de façon similaire — alors que les traitements ouverts étaient libres d'utilisation.

Il faut cependant veiller, ainsi que la première partie de l'étude le suggérait, à éviter qu'on ne présente comme de simples « aides à la décision » des systèmes qui sont de fait en position de trancher seuls. Pour cela, il sera fait obligation au responsable de traitement de décrire, au titre de ses obligations de transparence, « l'apport propre du décideur humain intervenant après l'automatisme, en particulier les méthodes et critères sur lesquels il fonde son appréciation ». S'il ne le fait pas, le responsable de traitement sera présumé fonder exclusivement sa décision sur le traitement automatisé. Il sera par conséquent frappé par l'interdiction de principe.

**25. Rôle des États en matière d'exceptions.** En l'état actuel du droit, le principe d'interdiction est assorti de plusieurs exceptions. Faut-il proposer une modification sur ce point ? Souvenons-nous de la décision rendue par le Conseil constitutionnel, examinant la loi qui adoptait le droit interne français au RGPD. Le Considérant 71 avait retenu l'attention de la doctrine, qui décidait : « (…) 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 »[43]. Pour le Conseil, certaines décisions sont si lourdes d'enjeux qu'elles ne tolèrent pas l'opacité. Le fait qu'une « intervention humaine » puisse être sollicitée dans un second temps n'y changera rien : il n'est pas acceptable qu'en première intention, la situation de la personne concernée soit fixée selon des mécanismes inaccessibles. Il faut en effet garder à l'esprit que la garantie tenant à une intervention humaine n'a pas vocation à être invoquée systématiquement par *tous* les destinataires de décisions individuelles — sans quoi l'automatisme n'aurait absolument servi à rien. Certains destins auraient donc été scellés sans qu'on sache comment ni pourquoi.

Certes, la réserve d'interprétation formulée par le Conseil est compatible avec le droit européen actuel. En effet, parmi les exceptions à l'interdiction de principe, aussi bien le « consentement de la personne concernée » que « la nécessité pour la formation ou l'exécution du contrat » sont inapplicables à la matière administrative[44]. La seule exception restante, le droit de l'État membre, est par définition même ouverte à une approche nationale spécifique. Mais ce raisonnement n'est pas transposable au droit privé, pour lequel les deux autres exceptions sont disponibles. Or, ne peut-on concevoir que dans certaines relations asymétriques de droit privé — droit du travail, de la

43 Décision n° 2018-765 DC du 12 juin 2018.

44 Les lignes directrices du G29 sur le consentement exposent en effet à juste titre que la relation entre administration et administré, structurellement déséquilibrée, est en principe incompatible avec l'expression d'un consentement ou d'un refus serein : Lignes directrices 5/2020 sur le consentement au sens du règlement (UE) 2016/679, 4 mai 2020, n° 16. Pour un commentaire de ce texte : N. Metallinos, « Lignes directrices du G29 sur le profilage. Tentative de clarification sur le régime spécial applicable au profilage et à la prise de décisions automatisée », CCE, 2018, n° 2, comm. 14.

consommation, des assurances — il faille considérer également que les décisions les plus lourdes d'enjeux sont insusceptibles d'être délivrées par une boîte noire ? Pour ces cas, si l'on veut éviter des discussions potentiellement difficiles sur l'applicabilité des exceptions — sur les qualités du consentement, sur l'absolue nécessité du recours à un automatisme pour la formation ou l'exécution du contrat — le plus simple reste de s'en tenir au principe d'interdiction.

Nous avons certes regretté, dans des écrits antérieurs, que le RGPD fasse la part trop belle aux marges d'interprétation nationale. Mais ces regrets portaient sur de *pures* questions de droit des données personnelles. Ainsi, il est absurde de pouvoir fixer l'âge de la « majorité informationnelle » des enfants entre 13 et 16 ans à la discrétion de chaque État membre[45]. En revanche, la question des décisions automatisées n'est pas de pur droit des données personnelles, mais est nécessairement mélangée de droit sectoriel national, selon le domaine dans lequel elle intervient.

Il nous paraît donc souhaitable que les exceptions de la « nécessité pour le contrat » et du « consentement » ne soient ouvertes au responsable de traitement qu'à défaut d'opposition du droit national. Le recours à des boîtes noires faisant émerger des décisions ayant des conséquences juridiques ou des impacts significatifs d'une autre nature sur les personnes ne serait pas un droit pour les responsables de traitement, mais relèverait bien plutôt d'une tolérance, qui pourrait leur être refusée.

**26. Garanties**. Supposons cependant que l'une des exceptions puisse être invoquée par le responsable de traitement sans que le droit national ne soit s'y soit opposé. Par exemple, une entreprise ayant publié une offre d'emploi reçoit des candidatures en si grand nombre qu'elle est contrainte d'opérer un « prétri » automatisé. Quelles sont les garanties applicables ?

**27. Droit à une motivation individuelle**. La première étape consiste ici encore à expliquer la décision qui vient d'être prise. S'agissant des traitements automatisés transparents, l'exercice était simple, puisqu'il s'agissait seulement d'exposer comment une formule-cadre avait été appliquée à des valeurs individuelles. Dans le cas d'un traitement opaque, le besoin d'éclaircissement apparaît bien plus fort, mais il est paradoxalement plus difficile à satisfaire. Si le traitement à l'œuvre est « autoapprenant » et que sa logique propre est impénétrable, l'exercice semble même, à première vue, impossible. Il ne l'est pas nécessairement. Il est tout à fait envisageable, par exemple, qu'un traitement algorithmique sophistiqué de sélection de CV recherche plusieurs qualités distinctes. Peut-être évalue-t-il d'une certaine façon l'adéquation de la formation initiale du candidat au poste proposé (par comparaison avec une base de mots-clés), d'une autre façon son expérience professionnelle (en pondérant différemment les anciens employeurs selon leur notoriété ou leur taille), et par un procédé encore distinct la qualité de son expression écrite telle qu'elle ressort d'une lettre de motivation (en s'appuyant sur l'apprentissage machine à partir de lettres étiquetées « bonnes » ou « mauvaises »). Dans cet exemple, peut-être est-il possible d'expliquer au candidat qu'en dépit d'une formation initiale jugée satisfaisante, l'expérience professionnelle a été considérée comme globalement insuffisante. Lorsque le traitement opaque est de type déductif, il sera d'autant plus facile de subdiviser les critères qu'on lui ordonne d'appliquer en grandes rubriques, afin de faire émerger une motivation.

45 Art. 8.1 du RGPD.

Aussi considérons-nous, à rebours de la solution actuelle, qu'il faut exiger du responsable de traitement qu'il accompagne la décision d'une motivation individuelle[46]. Celle-ci pourra être elle-même produite de manière entièrement automatisée, sans quoi le recours à un logiciel n'aura servi à rien. Cette motivation ne sera pas soumise à une exigence de qualité — on voit mal comment en juger — mais l'on peut compter sur une incitation structurelle du responsable de traitement à ce qu'elle soit la meilleure possible : une motivation peu convaincante constituera un aiguillon qui encouragera la personne concernée à solliciter ensuite « l'intervention humaine », autrement dit à former un recours.

À ceux qui estimeraient qu'une décision dont les ressorts intimes restent opaques est nécessairement une décision inexplicable, on proposera un nouveau parallèle avec la décision humaine. Un jury de recrutement comprend autant de « boîtes noires » que de membres. Chacun appréciera la lecture d'un CV ou le déroulement d'un entretien d'embauche à la lumière d'une longue expérience forgée à partir d'un corpus d'apprentissage. Il lui sera parfois impossible de savoir *très exactement* pourquoi un candidat lui a semblé « plus clair » ou « plus habile », mais il pourra rendre compte de ce diagnostic général.

S'il n'est vraiment pas possible de faire émerger quelque motivation individuelle que ce soit du traitement, alors le responsable devra assumer devant les destinataires de ces décisions qu'il est incapable de fournir un début d'explication à ses choix.

La motivation individuelle, même automatisée, viendra utilement compléter les exigences actuelles tenant à une information globale sur « la logique sous-jacente » du traitement, et la rendra bien plus concrète.

**28. Intervention humaine**. Si cette motivation individuelle, vraisemblablement automatisée, n'a pas convaincu, doit commencer la phase « d'intervention humaine ». Sans bouleverser les solutions actuellement en vigueur, il convient simplement d'en préciser le détail et la chronologie, par exemple de la manière suivante :

« Après avoir pris connaissance de la décision et de la motivation qui l'accompagne, la personne concernée a le droit de former un recours. Elle est alors invitée par le responsable de traitement à présenter des observations écrites au soutien de sa cause. Une nouvelle décision motivée est prise par un être humain, qui remplace entièrement la première. Les motivations de la décision humaine ne peuvent s'appuyer sur les résultats du traitement automatisé opaque ».

La dernière phrase doit être bien comprise. Il n'est évidemment pas interdit à l'être humain d'aboutir aux mêmes conclusions que la machine, mais il doit y parvenir par ses propres voies. Le diagnostic selon lequel la formation initiale du candidat n'était pas suffisamment en phase avec le poste proposé était peut-être excellent. Mais cela doit être démontré directement à partir du CV. « L'ordinateur l'avait dit » constitue éventuellement un indice pour l'investigateur humain, mais certainement pas un argument.

46 Sur la question de savoir si les textes actuels dirigent vers un droit à l'explication individuel ou collectif, ex post ou ex ante : J. Rochfeld, art. préc. in Rép. Dalloz IP/IT, n° 23 ; C. Castets-Renard, « Régulation des algorithmes et gouvernance du machine learning : vers une transparence et "explicabilité" des décisions algorithmiques ? », Revue Droit & Affaires, n° 15, déc. 2018, 1. Pour une conception plus collective du droit à l'intelligibilité du traitement : CNIL, Comment permettre à l'homme de garder la main ?… , rapport précité, p. 53.

Bien évidemment, si tous les destinataires de décisions automatisées demandent une révision humaine, alors le recours à la machine sera vain. C'est déjà le cas en droit positif. La motivation individuelle automatisée constitue une puissante incitation du responsable de traitement à avoir une idée, même approximative, des raisons qui justifient le classement des cas.

**29. Contrôle des discriminations**. Il reste à s'assurer que l'automatisme n'est pas source de discriminations prohibées. S'agissant des traitements transparents, la discrimination, par hypothèse, ne pouvait être qu'indirecte. S'agissant des traitements opaques, les discriminations directes sont possibles, puisque les critères fondant la décision demeurent dissimulés. Il ne s'agit pas d'un risque propre à l'ordinateur : rappelons une nouvelle fois qu'un jury humain est un ensemble de boîtes noires, qui peuvent dissimuler derrière une motivation de façade leurs motivations sexistes ou xénophobes. Il convient ici encore d'aménager la charge de la preuve : si une personne concernée démontre qu'une discrimination est vraisemblable, c'est au responsable de traitement d'apporter la preuve qu'elle n'a pas eu lieu — ce qui constitue une nouvelle incitation à maîtriser au moins les grandes lignes du fonctionnement de son traitement. La discussion sera nourrie par la production d'un rapport montrant si la présence du critère litigieux a un impact statistique sur le sens de la décision adoptée. On ne saurait trop conseiller aux responsables de traitement de procéder de leur propre chef à des contrôles *ex post* réguliers de ce type[47]. Cela leur permettrait de constater rapidement qu'un traitement produit des résultats anormaux, puis de rechercher les causes de ce comportement.

**

Refermons cette étude sur une proposition concrète de réécriture de l'article 22 du RGPD, qui tire profit de l'ensemble des observations précédentes et en réalise la synthèse.

## CONCLUSION

### Proposition d'ajout aux articles 13 et 14 du RGPD

X. Lorsque le traitement a pour finalité l'aide à la prise d'une décision produisant des effets juridiques à l'égard de la personne concernée par le traitement de données personnelles ou l'affectant significativement de manière similaire, le responsable de traitement décrit l'apport propre du décideur humain intervenant après l'automatisme, en particulier les méthodes et critères sur lesquels il fonde son appréciation.

### Proposition de réécriture de l'article 22 du RGPD

1. Le présent article s'applique aux décisions produisant des effets juridiques à l'égard de la personne concernée par le traitement de données personnelles ou l'affectant significativement de

47 Les lignes directrices précitées du CEPD fait figurer, au titre des « bonnes pratiques » (p. 36) : « des contrôles réguliers d'assurance qualité de leurs systèmes pour veiller à ce que les personnes soient traitées équitablement et ne fassent pas l'objet de discriminations, que ce soit sur la base de catégories particulières de données à caractère personnel ou autrement (...) ».

manière similaire. Sont considérés comme des décisions les processus ayant pour objet de choisir entre plusieurs actions ou abstentions possibles à l'égard de la personne concernée.

Chaque État membre peut établir une liste de critères qui ne peuvent faire l'objet d'aucune discrimination directe ni indirecte dans la prise de décisions à l'égard des personnes concernées.

2. Un traitement automatisé destiné à produire une décision individuelle est dit « transparent » lorsque la nature et la pondération des variables en entrée sont intégralement fixées et décrites, avant son exécution, dans une décision-cadre dont le contenu est accessible aux personnes concernées et dont les décisions individuelles constituent la simple exécution. Si la décision-cadre n'est pas directement formulée dans un langage simple et clair, elle est accompagnée d'une explication présentant ces qualités.

Il est communiqué à la personne concernée, à sa demande, l'état des variables employées pour produire la décision individuelle qui lui est opposée, de manière à ce qu'elle puisse vérifier que la décision-cadre lui a été appliquée sans erreur.

La contestation individuelle ou collective du contenu de la décision-cadre ne relève pas du présent texte.

Si une personne concernée ou un groupe de personnes concernées présente des éléments de fait laissant supposer que le traitement automatisé transparent est à l'origine de discriminations indirectes fondées sur un critère visé au 1. du présent article, il incombe au responsable de prouver que sa décision n'est pas fondée sur un tel critère, après avoir produit un rapport statistique permettant de vérifier s'il existe une corrélation entre la présence du critère litigieux et le sens de la décision.

3. Un traitement automatisé destiné à produire une décision individuelle est dit « opaque » lorsque le responsable de traitement prive la personne concernée d'une description exhaustive des règles qui lui sont appliquées, que ce soit par choix ou en conséquence des techniques employées par le traitement.

4. Il est interdit de fonder exclusivement sur un traitement automatisé opaque une décision produisant, pour la personne concernée, des effets juridiques ou l'affectant significativement de manière similaire.

Un responsable de traitement qui affirme recourir à de simples systèmes d'aide à la décision est néanmoins réputé fonder exclusivement sa décision sur un traitement automatisé si, dans l'exécution de ses obligations de transparence issues des articles 13 et 14 du présent règlement, il n'a pas indiqué quel était l'apport propre du décideur humain intervenant après l'automatisme, en particulier les méthodes et critères sur lesquels il fonde son appréciation.

5. Par exception au paragraphe 3, la décision peut être autorisée par le droit de l'Union ou le droit de l'État membre auquel le responsable du traitement est soumis et qui prévoit également des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée.

6. Par exception au paragraphe 3, la décision est autorisée dans les deux cas suivants, sauf si le droit de l'État membre auquel le responsable du traitement est soumis s'y oppose. La décision :

a) est nécessaire à la conclusion ou à l'exécution d'un contrat entre la personne concernée et un responsable du traitement ;

ou ;

b) est fondée sur le consentement libre et explicite de la personne concernée.

7. Dans les cas visés au paragraphe 6, la personne concernée bénéficie au moins des garanties suivantes :

a) la décision est accompagnée d'une motivation individuelle, qui peut être elle-même produite exclusivement par un traitement automatisé ;

b) après avoir pris connaissance de la décision et de la motivation qui l'accompagne, la personne concernée a le droit de former un recours. Elle est alors invitée par le responsable de traitement à présenter des observations écrites au soutien de sa cause. Une nouvelle décision motivée est prise par un être humain, qui remplace entièrement la première. Les motivations de la décision humaine ne peuvent s'appuyer sur les résultats du traitement automatisé opaque ;

c) si une personne concernée ou un groupe de personnes concernées présente des éléments de fait laissant supposer que le traitement automatisé opaque est à l'origine de discriminations directes ou indirectes fondées sur un critère visé au 1. du présent article, il incombe au responsable de prouver que sa décision n'est pas fondée sur un tel critère, après avoir produit un rapport statistique permettant d'observer s'il existe une corrélation entre la présence du critère litigieux et le sens de la décision.