



**HAL**  
open science

# The human and machine part in "automated" decisions. Proposals for a rewriting of article 22 of the GDPR

Emmanuel Netter

## ► To cite this version:

Emmanuel Netter. The human and machine part in "automated" decisions. Proposals for a rewriting of article 22 of the GDPR: Translation of an article to be published in french in Vanessa Barbé, Les algorithmes et le droit, Mare et Martin. 2022. hal-03701045

**HAL Id: hal-03701045**

**<https://hal.science/hal-03701045>**

Preprint submitted on 21 Jun 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 4.0 International License

## The human and machine part in "automated" decisions. Proposals for a rewriting of article 22 of the GDPR.

Emmanuel Netter  
Professor of private law at the University of Avignon  
contact @ enetter . fr

Translation by the author of [an article originally in French](#), in the forthcoming book *Algorithms and the Law*, dir. V. Barbé and S. Mauclair, Mare et Martin

*"The decision we have made is open to criticism, everyone is entitled to a feeling. To confuse that feeling with a judgment is something else. To arrive at this judgment, the court read 4,800 pages of proceedings, spent twenty-six hours in debates, listened to 15 lawyers and a prosecutor, and met as a collegium for several hours. This does not mean that our decision is the right one, but what makes this decision a judgment is that we each kept our emotions and impulses at bay.*

**Définir une variable para\_noye = para\_noye + 1. Decide well.** What is a good decision? The answer reproduced above is that of Michaël Humbert, president of the correctional court that ruled in the "Mila" case, named after this young woman who had been the subject of fierce attacks on social networks because of her positions on Islam<sup>1</sup>. The type of decision in question here is therefore one of the most solemn, one of the most high-stakes that can be conceived: should the wrath of the criminal justice system be hurled at a defendant, and with what power? Before reaching its decision, the Tribunal heard the lawyers and representatives of the public prosecutor's office, exchanged views with them and debated within the Tribunal. Discussions, time and doubts: all elements that seem characteristic of human decision-making processes.

**Définir une variable para\_noye = para\_noye + 1. Deciding with the machine.** Two years ago, Estonia announced, with the marketing sense that characterizes it when it comes to digital technologies, that it intended to delegate to computer systems the task of making decisions in criminal matters<sup>2</sup>. It was also specified that any dissatisfied defendant would have the right to have a human judge review the entire case. It is enough to imagine, which Estonia did not specify in its communication to the general public, that the public prosecutor's office would undoubtedly have a symmetrical right in the event of a decision that was too lenient, to deflate the innovation somewhat and to agree with Antoine Garapon's observation that there is currently an element of "bluffing" around automated decision-making systems<sup>3</sup>.

Florence G'Sell, however, invited us not to underestimate the appetite of litigants for a "faster and above all simpler" judicial system<sup>4</sup>. It is true that computerized processing is characterized by its speed, but this is not its only attraction. Let us return to the qualities presented by Mr. Humbert as contributing to a quality decision. Among them is the ability to take into account exhaustively thousands of pages of procedure, tens of hours of debate, a performance at the limit of the human mind's capacity. However, multiply this volume of information by ten, a hundred or a thousand: the machine will not encounter any difficulty. As for the ability, mentioned by the president of the court,

---

1 H. Seckel, "'What you don't do in the street, don't do on social networks' : de quatre à six mois avec sursis et une leçon de conduite pour les harceurs de Mila ", article lemonde.fr of July 7, 2021.

2 See, e.g., H. Gardette, "Will robots make good judges?", France Culture broadcast, May 23, 2019, with F. G'Sell, B. Jean, and E. Polnas.

3 "Mistrust on "robot-judges" in the courts", article 24heures.ch of April 26, 2019.

4 The above-mentioned program "Will robots make good judges?"

to "keep [one's] emotions, [one's] impulses at bay", it is obviously a natural attribute of computing systems.

Computers therefore have a number of assets that can be useful for good decision-making<sup>5</sup>. There is nothing surprising in this: to decide is to take one of the paths open, abandoning the others, on the basis of the information available; and computer science is the science of information processing.

**Définir une variable para\_noye = para\_noye + 1. Who holds the tool?.** The capabilities of the machines must be put to good use. Cathy O'Neil's famous book, *Weapons of Math Destruction*, has largely contributed to drawing attention to the misuse of decisional algorithmic processing<sup>6</sup>. In particular, the author tells the story of an American teacher who was fired despite the high opinion she was held in by students, parents and colleagues. The "algorithm" had in fact given her a score that placed her among the poorest teachers in her region. How the indicator was constructed was not revealed; the recipient of the decision was not able to comment. An opaque and arbitrary decision-making process became unquestionable because it had been painted in the colors of science. Objectivity was presented as certain; the fairness of the result was irrefutably presumed.

The story told by Cathy O'Neil is certainly uplifting and terrifying. But it is the result of a fundamental misunderstanding: there is no reason to bow down to the computer as to a pagan deity. It is like worshipping a screwdriver or a pair of scissors. Let us rather ask ourselves who is mobilizing these tools, and what they are trying to do. The irruption of information technology into a decision-making process is neither good nor bad news in itself. It all depends on how it is used, and how it is articulated with the human will.

**Définir une variable para\_noye = para\_noye + 1. The computer never decides alone.** Any so-called "computer decision" encapsulates a human decision. Rather than opposing one to the other as if they were strangers, it is better to observe how they interpenetrate.

Let us consider a first family of computer processing, which we will call "deductive"<sup>7</sup>. The human provides the machine with a set of ready-to-use rules, which it only has to execute. At this stage, there is no way to prejudge the content of these rules, either favorably or unfavorably. In order to determine the performance of a set of employees, one can imagine assigning to each letter of their last name a number corresponding to its rank in the alphabetical order, adding the whole, dividing it by the number of their birth department and finally multiplying it by the current temperature in degrees Celsius. The resulting score may be perfectly absurd, but if no one is allowed access to the way it is constructed, it will be difficult to dispute. But we can imagine another example. For the 100 meter race of the Olympic Games, I can ask a machine to start the race with a sound signal, to stop the stopwatch of each runner when it will be detected that he/she crosses the finish line, and then to display the result on the stadium board in ascending order of race time. The process is possibly at the mercy of technical incidents, but it is perfectly sound in its general conception. Let's remember that, in this first family of automatisms, the human fixes directly and *a priori* all the applicable rules.

---

5 On this question and the whole theme of this study: S. Desmoulin-Canselier and 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. The book has been translated into French as *Algorithmes: la bombe à retardement*, Les Arènes, 2018.

7 For a presentation of the different types of algorithmic processing, including those that we describe below as "inductive", see, for example, S. Fischman and 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, spec. I.

The situation is different in the family of treatments that we will call "inductive", in that they are supposed to examine concrete cases, experimental data, and then to derive from them the rules that should be applied in order to arrive at a satisfactory result, which is called "automatic learning" or "machine learning"<sup>8</sup>. This is how "supervised learning" works: the human being provides the computer with a training corpus, made up of training data that he himself has assigned to a category. Here are 100,000 pictures of cats, as many pictures of dogs, and it's up to you to discover and formulate the relevant criteria to differentiate these two animals. It could also be *resumes* of job applicants whose careers have been either successful or unsuccessful, or profiles of bank borrowers who have paid their bills or defaulted on their loans. Machine learning can also take an unsupervised form, guided by "reinforcement". The human being intervenes here - as we said, he *always* intervenes - by assigning an objective. It may be to make a robot, that does not yet know how to move, climb stairs, to learn to become the best Go player in the world or to minimize the time that employees working in a skyscraper spend waiting for elevators. The system will experiment with different approaches, and will be able to measure its own performance against the initial goal.

**Définir une variable para\_noye = para\_noye + 1. The "black boxes"**. It follows from the above that computer systems aiming at making decisions are intrinsically neither objective nor arbitrary, but that they are tools handled by the human will. Moreover, it appears that the logic governing these algorithms can remain inaccessible to the recipients of the decisions. In the case of "deductive" processing, this opacity will always result from a choice made by the data controller since, by definition, the rules applied are known *ex ante*. He is therefore in a position to reveal them, but he may be tempted to encapsulate them in a technical process that does not let them appear, only the decision being revealed. On the other hand, certain "inductive" processing operations are opaque not only for the recipients of the decisions, but also for the controller himself. Indeed, it happens that the machine, trained from a qualified corpus or pursuing an objective by reinforcement, produces for itself the set of rules that it assumes to be the most efficient, in a form that is no longer accessible to human intelligence<sup>9</sup>. This *erga omnes* opacity is then without remedy.

The dangers of these "algorithmic black boxes" have been denounced many times, and rightly so<sup>10</sup>. But once again, the point needs to be qualified. Opaque decision-making processes did not appear with computers. A few years ago, men were sent to prison, and before that to the scaffold, on the "intimate conviction" of a jury<sup>11</sup>. For a long time, this seemed acceptable because a rigorous procedure governed the presentation and discussion of evidence, and the collegiality of the debates mixed individual subjectivities in order to sublimate them into a supposed collective objectivity. It

8 These are the techniques that are most often asserted to be part of "artificial intelligence". See however, for a much more encompassing definition of AI, S. Merabet, *Vers un droit de l'intelligence artificielle*, pref. H. Barbier, Dalloz, Nouvelle bibliothèque de thèses n° 197, 2020. This study will not use the concept of AI, which is not currently a consensus in either law or computer science. However, see below the existence of a draft European regulation on the subject.

9 See in particular. 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 E.g. E. Mouriessé, "L'opacité des algorithmes et la transparence administrative", RFDA, 2019, p. 45; T. Douville, "Parcoursup 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*, 2017 report, p. 53.

11 The requirement to state the reasons for the principle of conviction by the jury of assizes was introduced in article 365-1 of the Code of Criminal Procedure by a law n° 2011-939 of August 10, 2011. A decision of the Constitutional Council No. 2017-694 QPC of March 2, 2018, subsequently required that the reasoning also cover the quantum of the sentence.

is possible to recruit or promote an employee, or to choose a tenant without having to explain it, but favorable rules of evidence will come to the rescue of anyone who feels he or she has been discriminated against<sup>12</sup>.

### **Définir une variable para\_noye = para\_noye + 1. Presentation of the current framework.**

Basically, what we expect from a decision-making process, whether it is automated in whole, in part or not at all, is that it presents certain *qualities* (of fairness, of justice) that we hope to obtain through the implementation of *guarantees*, which vary according to the context (transparency of the procedure, motivation of the choice, existence of appeals...)<sup>13</sup>.

The shortcomings of "purely human" decisions have been identified for a long time, and quality guarantees have been adjusted accordingly, in each sector considered (labor, examinations and competitions, justice, housing, banking, etc.). On the other hand, the relatively rapid rise in power of computerized decision-making or decision-support systems has encouraged the European legislator to ask, in a transverse manner, whether these systems present their own dangers, and whether it is appropriate to enact specific guarantees for them<sup>14</sup>. The GDPR offered a coherent framework for this, since the IT procedures aimed at establishing the rights and obligations of a given individual necessarily have personal data as their fuel. The wealth of individual identity is suddenly reduced to a few indicators considered relevant by the data controller, so that a decision can be made: this is indeed data processing, subject to the Regulation<sup>15</sup>.

From this comes Article 22 of the GDPR, which states first: "*1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.*" "Profiling" is understood to mean the assignment of the person whose data is processed to a category, such as "reliable debtors" or "poor performers", which may be the prerequisite for a decision (for example, to grant credit or refuse a promotion)<sup>16</sup>.

The text hits hard, but on a relatively narrow target. The target is narrow, because only decisions based "solely" on automated processing are concerned. A strict reading of the provision implies that whenever the decision is based on at least one element other than automated processing, it falls entirely outside the scope of the text, even if the contribution of automation is preponderant, or even overwhelming, in relation to the other sources mobilized. But when the entire decision-making process is based on algorithmic processing, the text is very strong: even if exceptions are provided for, the principle is the prohibition.

---

12 See in particular Art. L. 1132-3-3 of the french Labour law code.

13 With regard to computerized decisions, S. Desmoulin-Canselier and D. Le Métayer, op. cit. look for the following qualities in each field studied: "legitimacy", "efficiency, reliability", (absence of) "discrimination, bias", "control, intelligibility", "protection of personal data and privacy".

14 On this question as a whole, see in particular, in addition to the above-mentioned work by S. Desmoulin-Canselier and D. Le Métayer, spec. p. 123 ff. 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 GDPR*, ed. N. Martial-Braz and 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*, May 2020; T. Douville, *Droit des données à caractère personnel*, Gualino, 2021, n° 507 s.

15 Automated decision-making could also be regulated by the proposed European regulation on AI, assuming it is adopted one day (*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*). AI is indeed defined in article 3.1 as "*software 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 On this notion: N. Martial-Braz, "Le profilage. Fiche pratique," CCE, 2018, n 4, file 15.

In French law, Article 47 of the *loi informatique et libertés* (French Data Protection Act) states:

*"No judicial decision involving an assessment of a person's conduct may be based on automated processing of personal data intended to evaluate certain aspects of that person's personality.*

*No decision having legal effects on a person or significantly affecting him or her may be taken solely on the basis of automated processing of personal data, including profiling (...)"<sup>17</sup>.*

The French approach to the question thus presents several specificities. To begin with, a first paragraph that sets out a specific prohibition on judicial decisions, contrary to the Estonian position. Once a treatment is used to "evaluate certain aspects of the personality", it cannot be used as a basis for a decision, even partially<sup>18</sup>. Then, in the second paragraph, we leave behind the specific field of court decisions to encounter a transcription of the general solution of Article 22 of the GDPR. But it is not expressed, as in the Regulation, in the form of a subjective right of the data subject "not to be subject" to certain types of decisions<sup>19</sup>.

Let us return to the European text. There are three exceptions to the prohibition in principle<sup>20</sup>. A decision based exclusively on automated processing may be authorized by Union law or the national law of the controller, subject to "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests" (b). This leeway has been used in particular by France in the area of administrative decisions<sup>21</sup>. The decision is also permitted if it is "necessary for entering into, or performance of, a contract between the data subject and a data controller" (a) or if it is based on the "on the data subject's explicit consent" (c). Whereas exception (b) left it entirely to European and national law to provide adequate safeguards, cases (a) and (c) trigger the application of a core protection<sup>22</sup>. Indeed, "(...) the controller shall implement appropriate measures to safeguard the rights and freedoms and legitimate interests of the data subject, **at least** the right of the data subject to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision".

Article 22 of the GDPR concludes with specific provisions for processing "revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the

17 « *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 (...)* ».

18 L. Huttner, "Données à caractère personnel - Décision automatisée et justice", Répertoire Dalloz IP/IT et Communication, n° 14, notes that this French position results in a stricter framework for automated decisions than that of the GDPR, on an issue for which the regulation does not seem to provide any national leeway. The question of compliance with European law is therefore raised.

19 On the difference between the consecration of a subjective right and the enactment of a prohibition: J. Rochfeld, art. préc. in Rép. Dalloz IP/IT, n° 13. T. Douville, op. cit. at n° 573.

20 Art. 22, 2° of the GDPR.

21 Article 47 of the *Loi informatique et libertés* provides for the case of "(...) 2° Individual administrative decisions taken in compliance with Article L. 311-3-1 and Chapter I of Title I of Book IV of the Code of Relations between the Public and the Administration, provided that the processing does not concern the data mentioned in Article 6 of this Act. These decisions shall include, under penalty of nullity, the explicit mention provided for in Article L. 311-3-1 of the Code of relations between the public and the administration. For these decisions, the data controller shall ensure that the algorithmic processing and its evolution are under control, so as to be able to explain, in detail and in an intelligible form, to the data subject the way in which the processing has been implemented with regard to him or her. As the rest of the study focuses on improving the texts at the European level, this specific French regulation will not be developed.

22 Art. 22, 3° of the GDPR.

processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation"<sup>23</sup>. For them to constitute the exclusive basis for a decision producing legal effects, or significantly affecting the person in a similar way, additional conditions are required<sup>24</sup>.

Automated decisions are also subject to a framework provided by Articles 13 and 14 of the Regulation. These texts require that data subjects be given a certain amount of information on the identity of the controller, the purpose of the processing, the storage periods, etc. However, they also require notification of "the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject"<sup>25</sup>.

**Définir une variable para\_noye = para\_noye + 1. Criticism and proposals.** This European framework for automated decisions is very welcome in principle. But is it satisfactory in its modalities? We will try to show that the current wording of article 22 is open to criticism on several points (I) and we will make proposals for improvement (II).

## I - A critique of Article 22 of the GDPR

The text of Article 22 appears to be perfectible both from the point of view of the decisions referred to (A) and of the guarantees deployed when it is allowed, by exception, for a decision to be taken exclusively on the basis of automated processing (B).

### A - Inadequate criteria

The Regulation refers to "decisions" producing legal effects or similarly affecting the data subject taken "solely on the basis" of "automated processing". The concept of "decision" requires simple clarification (1). On the other hand, the criterion of "automatic processing" is too broad, imposing a ban in principle on processing that does not deserve such a challenge (2). The criterion of "exclusive basis" is too narrow, which allows potentially dangerous automated processes to escape the text (3).

#### 1 - The "decision": a concept to be clarified

**Définir une variable para\_noye = para\_noye + 1. Presentation.** Some elements of the scope of Article 22 seem to need to be rethought. This may not be the case for the notion of "decision", which calls for only a few interpretative remarks. It does not operate independently: the regulation

---

<sup>23</sup> Art. 22, 4° of the GDPR.

<sup>24</sup> In addition to falling within the scope of one of the three exceptions mentioned above (authorization by Union or national law, necessity for the performance of a contract, or explicit consent of the individual), Article 9.2 (a) or (g) must also be invoked. Article 9.2 (a) provides that "the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject". Article 9.2, g) states: "processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject".

<sup>25</sup> Articles 13, 2, f) and 14, 2, g) of the GDPR.

refers to decisions "having legal effects" on the data subject or "affecting him or her significantly in a similar way".

**Définir une variable para\_noye = para\_noye + 1. Extensive interpretation: the splittability of decisions.** Let us first observe that it may be necessary, in order to keep the text as effective as possible, to split a seemingly global position of the controller into several small "decisions". Imagine a bank that uses human arbitration to determine to whom it grants or refuses credit, but then allows an automated system to set the interest rate for accepted transactions, without supervision. To stick to a global analysis would suggest that a "decision" to lend at 1.25% was made jointly by the human and the machine. However, we must restore the exact qualification of the set-up: two decisions were made, one on the principle and the other on the rate. The second was therefore entirely automated: we shall see later what legal regime this observation calls for.

**Définir une variable para\_noye = para\_noye + 1. Restrictive interpretation: a procedure to fix an individual situation.** This means that in a given situation, one must be careful to find as many decisions as necessary. This does not mean that one should see decisions everywhere. Having given the concept some impetus, let's try to keep it in proper proportion. Let's imagine a merchant whose business appears in a very favorable position in Google Search results following a very popular query (such as the words "used car" followed by the name of a city). Now the Mountain View firm revises its algorithmic treatment, as it does regularly: the merchant is suddenly relegated to the second page of the results, its visibility collapses and its sales with it. Can we consider that the arrangement of hyperlinks in response to a search constitutes a decision? One could try to get around the difficulty by relying on the fact that it is not, in any case, a decision "producing legal effects". In the context of "natural" results, as opposed to "sponsored" results, the search engine is not in a contractual relationship with the person concerned, and the hypothesis studied does not engage its extra-contractual civil liability. But what about decisions "affecting (the person) significantly in a similar way"? Is this not the case here? Recital 71 of the GDPR cites as examples "the automatic rejection of an online credit application or online recruitment practices without any human intervention". Denying individuals the benefit of a contract they seek to enter into, these decisions could practically be considered to have legal effect. Decisions affecting the individual "in a similar way" would then be quasi-legal or para-legal decisions. But the EDPB offers a broader interpretation, considering that "the level of significance must be similar to that of a decision with legal effect"<sup>26</sup>. Such a decision, he adds, "must be of such a nature as to significantly affect the situation, behaviour or choices of individuals" and "have a prolonged or permanent impact on the person concerned". This could be the case, it seems to us, of the relegation of a company to the second page of Google search results<sup>27</sup>. The scope of application of Article 22 thus conceived would, however, appear too broad. But the definition of "decision" can be narrowed to avoid this. Let us assume that the *purpose of* a decision is to maintain or alter the situation of the person concerned. Algorithmic treatments producing *simple incidental effects* on individuals, as in our example, are then excluded. As far as they are concerned, other bodies of rules can intervene in case

---

26 EDPB, Guidelines on automated individual decision making and profiling for the purposes of Regulation (EU) 2016/679, 2018, p. 24.

27 It should also be noted that in French law, Article 47 of the Data Protection Act refers to a decision "producing legal effects with regard to a person or affecting him or her in a significant way", without using the phrase "in a similar way". Comp. T. Douville, op. cit. at no. 581: "The decision must affect the person concerned, i.e. produce effects similar to legal effects".



the automatism causes an illicit damage - notably extra-contractual civil liability - but Article 22 of the GDPR does not constitute an adequate tool. As an example, on the other hand, the use of automated moderation tools triggering the removal of content on social networks, or even a simple degradation of its visibility, would indeed constitute a "decision": the process is intended to produce an individual verdict<sup>28</sup>.

The concept of "decision" therefore required some clarification, but does not in itself give rise to any insurmountable difficulties. The same cannot be said of the term "automated processing".

## 2 - "Automated processing": too broad a concept

**Définir une variable para\_noye = para\_noye + 1. The distrust by default of any fully automated process.** Suppose that a decision is to be made that produces legal effects or affects the individual in a similar significant way. There is then a prohibition - even if it is formulated as a subjective right of the person concerned - to base it entirely on "automated processing". There are certainly exceptions, which are followed by a series of specific guarantees. But the principle is indeed prohibition. Two corollaries follow from this. First, leaving the last word to automation is an evil in itself, and should be tolerated only in rare cases and with great precautions. Secondly: conversely, placing a human agent at the end of the decision-making chain is necessarily virtuous, since then Article 22 has nothing more to say.

Because they are too general, these statements are erroneous and may, in some cases, produce perverse effects.

They are quite understandable if we assume that the European legislator, in enacting these rules, was essentially aiming at inductive algorithmic processing, those that are likely to generate their own norms, for example by machine learning from qualified learning data. As Jean Foyer stated as early as 1977, as rapporteur of the Loi informatique et libertés for the National Assembly, "the computer has, it is said, no 'faculty of astonishment' in the face of errors of law or of fact, which may affect the data. Nor does it have any when faced with false deductions attributable to the program"<sup>29</sup>. If an automatic program is to learn to distinguish cats from dogs, but there are an overwhelming number of light-colored cats and dark-colored dogs in its learning corpus, it might well conclude that a light-colored animal is a cat. More ominously, if a treatment tasked with identifying promising resumes learns from a corpus in which "successful" career paths are predominantly those of white males, because racism and misogyny have artificially promoted these profiles over others over the past decades, the computer will incorporate this data - this is what happened to Amazon<sup>30</sup>. It is therefore essential to be wary of such inductive algorithms and to avoid

---

28 It is also a decision with legal effects, in two ways: it is the sanction of the norm violated by the user (general conditions, or state rule) and it affects a fundamental freedom (freedom of expression). It is therefore not surprising that the draft regulation on an internal market for digital services of 15 December 2020 provides that the use of automated moderation tools must be notified to the user (articles 14.6, 15.2 [c]) and, above all, that it prohibits that a decision in this matter "be taken solely by automated means" (article 17.5).

29 Full report of the session of the National Assembly of Tuesday, October 4, 1977, OJ, p. 5782.

30 J. Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," reuters.com article, October 11, 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."

letting them have the last word: in such cases, the intervention of a human "faculty of wonder" in the final analysis is necessary, and the provision of Article 22 is absolutely justified.

**Définir une variable para\_noye = para\_noye + 1. The harmlessness of some fully automated processes.** But let us now consider a purely deductive algorithmic processing, which would be satisfied with mechanically implementing criteria deliberated in advance according to human decision-making processes: the machine does not create any danger of its own. To attack the computer here is to shoot the messenger.

Take, for example, the algorithms published by the city of Nantes, one of which is used to calculate social pricing for transport, the other for social pricing for water<sup>31</sup>. The rules governing them have been adopted by the competent political bodies in the required form. The content is published, so that everyone can check that the translation of the language of the law into the language of machines is absolutely accurate<sup>32</sup>. For each citizen, the processing is able to produce a decision without any further human intervention. It is indeed a decision in the sense in which we understand it, namely a procedure having for object to carry out a choice intended to maintain or to modify the situation of an individual. This decision produces legal effects, since it modulates the amount of the claims that the city has against the water consumer.

It should also be noted that a human civil servant could apply the formula himself, with a pen and paper: the waste of resources would be phenomenal, without the quality of the results being likely to improve in any way<sup>33</sup>. The same will be true for almost all public and private billing systems, which will apply a regulatory or contractual scale by a purely automated means in order to make an individual decision. A similar logic applies to the marking of an examination in the form of a MCQ, for which an official scale has been adopted.

In these situations, the provisions of Article 22 appear inadequate. The text encourages the decision-maker to introduce human subjectivity after the passage of the automatism, in order to escape its scope. But if the common rule is right, it is not only inappropriate, but even dangerous, for a human agent to override it for an individual case, and this will in any case be legally impossible in most of the situations covered. If the common rule is problematic, it must be challenged by the appropriate legal means, but the difficulty does not stem from the presence of an algorithmic treatment, which is here only a faithful executor, a more sophisticated avatar of the abacus or the calculator.

The notion of "automated processing" therefore appears excessively broad. On the contrary, the notion of "exclusive basis" of the decision is too narrow.

### **3 - Decisions taken "exclusively on the basis" of automated processing: too narrow a criterion**

**Définir une variable para\_noye = para\_noye + 1. Decision support systems.** Let's focus again on automatic information processing systems based on unpredictable and potentially unknowable

31 [https://data.nantesmetropole.fr/pages/algorithmes\\_nantes\\_metropole/](https://data.nantesmetropole.fr/pages/algorithmes_nantes_metropole/).

32 On the transcription of an algorithm described by a legal text into computer code: Liane Huttner, Denis Merigoux. "Translating the law into code using programming language. Artificial Intelligence and Public Finance," Oct 2020, Nice, France, hal-03128248.

33 It should also be noted that such manual processing can still be considered "automated", depending on the meaning of the term. In the Larousse online, the first definition for "automatism" is certainly "Quality of devices and installations that function without human intervention", but the second is "Act accomplished mechanically, without intervention of the will".

criteria, for example those that exploit certain machine learning techniques. Such tools can be extremely valuable. For example, medical imaging analysis software can detect anomalies so subtle that they escape the human eye. The goal here is to get the doctor's attention, so that he or she can take control of the diagnosis. It is he, and he alone, who will make the decision to retain a therapy. Were one to have the crazy idea of letting the automatism itself opt for a medicinal or surgical treatment, one would come up against the prohibition in principle of basing this decision - which does not produce legal effects, but affects the person in a significant way - on an automated treatment. In other words, while Article 22 of the GDPR provides a framework for fully automated decisions, it does not prevent simple "decision support systems" (DSS).

**Définir une variable para\_noye = para\_noye + 1. The danger: automated decisions disguised as DSS.** But the danger is immediately apparent. Take an automatism which, the moment before, was producing a chain of authentic decisions. Now display the information it produces on a screen, in front of which you place a subordinate who will glance at it distractedly before systematically pressing a green button: are you faced with a simple "monitoring" tool that escapes article 22<sup>34</sup> ?

The EDPB has understood this and has adopted a more demanding interpretation of the text. It is only when "a human being considers and takes into account other factors in making the final decision" that we are dealing with a simple DSS<sup>35</sup>. "For human intervention to occur, the controller must ensure that any control over the decision is meaningful and not merely a token gesture. The control should be performed by someone who has the authority and competence to change the decision. All relevant data should be taken into account in the analysis.

However, there is an inherent weakness in the overall logic of "compliance" into which personal data law has tipped since the GDPR came into force: the deception may only be revealed at the time of one of the (rare) CNIL inspections. Admittedly, the EDPB considers that "As part of its data protection impact assessment, the controller should identify and record the degree of human intervention in the decision-making process and the stage at which this occurs." But the content of this impact assessment, which is not published, will remain hidden from the eyes of data subjects<sup>36</sup>.

**Définir une variable para\_noye = para\_noye + 1. Obligation to justify the human decision-maker's own contribution.** It seems possible to proceed otherwise. A simple decision aid, although in principle excluded from the scope of Article 22, is nonetheless a processing of personal data subject to the rest of the GDPR. However, the introduction to this study recalled the existence of Articles 13 and 14, relating to transparency obligations. When the data controller makes automated decisions, he must inform the data subjects and enlighten them on the "underlying logic" of the processing as well as on its consequences. These texts could be amended to create a specific information obligation for processing operations that are alleged to be simple decision-making tools<sup>37</sup>. This information would relate to the human decision-maker's own contribution, in particular to the criteria and methods he or she uses.

---

34 On this question, see in particular. J. Rochfeld, art. préc. in Rép. Dalloz IP/IT, n° 7 s.

35 Guidelines, supra, at 23.

36 Under Article 36 of the GDPR, even the supervisory authority will only be aware of the DPIA if the controller believes, after its analysis, that the risks to individuals are high. The risk that he will be lenient with himself is not zero.

37 In this sense: T. Douville, op. cit. at n° 581.

In the absence of such precision, the so-called decision support system would be presumed to operate exclusively on the basis of automated processing.

Assuming that the controller complies with this information obligation, but in a particularly elliptical manner, there would be grounds for alerting the data subjects or the supervisory authority, where they may currently be kept in complete ignorance.

As for the data controller who would have indicated a specific human contribution that would not correspond to the reality of his practices, he would be liable to the heavy administrative sanctions enacted by the GDPR in case of non-respect of the rights of the data subjects.

Having proposed several improvements to the criteria for the scope of application of Article 22, let us now examine the safeguards for data subjects when, by way of exception, a purely automated decision is allowed.

## **B - Unclear guarantees**

**Définir une variable para\_noye = para\_noye + 1. The triptych of guarantees.** The prohibition of taking decisions with high stakes exclusively on the basis of automated processing is only a principle, which is subject to exceptions. Let's leave aside those that would be provided for by the law of the Union or of the Member State with guarantees that we only know must be "appropriate". With regard to the other two exceptions - the need for the conclusion or performance of a contract and the explicit consent of the data subject - the GDPR does specify a set of guarantees: "at least the right of the data subject to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision".

The formula lacks clarity. It is articulated around three elements. Is their order meaningful? Are they steps in an overall process, arranged in chronological order? But then, what is the "human intervention" that takes place even before the person has expressed their views? Does expressing one's views necessarily support a challenge to the decision? If not, what is the point of expressing a critical view if the review is not sought?

Recital 71 of the Regulation is hardly more enlightening: "In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision". The notion of "specific information" is not spelled out, and the EDPB guidelines do not venture to define it. However, the contribution of this recital is to affirm that the data subject has the right to an *explanation of the* decision taken against him or her.

**Définir une variable para\_noye = para\_noye + 1. The components of "human intervention".** Everything finally seems clearer if we consider "human intervention" not as one of the guarantees in the list, but as the keystone of the whole. Let's say a person has just been notified of a decision made by a machine acting alone. The individual considers that he has been treated unfairly. The first thing to do is to place a human interlocutor in front of him, because one does not negotiate with a computer. To use a formula of the engineer Benjamin Bayart, "the computer is fatal"<sup>38</sup>.

---

38 See, e.g., in B. Bayart and M. Rees, "Health Pass, Geopolitics of Data, Private Copying?", Thinkerview Conference, June 15, 2021, <https://www.youtube.com/watch?v=EOWeewlc2CE>.

A wide range of interactions is then possible between the two people thus brought into contact. In a coherent chronological order, one could first envisage that the human agent of the data controller answers questions, explains the decision, and teaches. But is it always possible? The whole point of certain automated decisions, as we have said, is that they are produced in relative autonomy by the machine, according to a logic whose details remain inaccessible to us. Then, the explanation - or the absence of explanation - having eventually failed to convince the recipient of the decision, the latter can ask for a revision. It will then be up to the human being to reopen the investigation, as it were. It should be noted in passing that if a first agent has been charged with explaining - and thus, indirectly, defending - the automated decision, he or she is probably no longer in a sufficiently impartial position to rule on an appeal. Article 22 expects this second decision to be contradictory, i.e. the person concerned must have had the opportunity to present, if he or she so wishes, elements in support of his or her case.

Suggestions for improvement have been made both in terms of the scope of Article 22 and the guarantees offered to data subjects when a fully automated decision is made. Let us now see how these suggestions can be used to reorganize or even rewrite part of the text.

## II - A recast of Article 22 of the GDPR

It is proposed that two very different legal regimes be established, depending on whether the automated processing behind the decision is "transparent" (A) or "opaque" (B). The first is a favourable regime, the second a cautionary regime.

### A - The favourable regime for transparent decisional treatments

**Définir une variable para\_noye = para\_noye + 1. An authorization in principle.** In the first part of the paper, we used the example of the social water pricing of the city of Nantes to show the lack of danger of a purely executive automated processing, which only applies a framework decision at the individual level.

The current approach triggers the prohibition in principle and forces one to take refuge behind one of the exceptions provided for in Article 22, 2°. It is certainly possible to consider that such automatisms are based on national law when they are in the public sector, or on necessity for the performance of the contract when they are in the private sector. But it should not be necessary for a commercial company that wishes to apply its public tariffs mechanically to a given client to justify that this is *strictly necessary* for the performance of an agreement. To require this would be to water down the exception, which will have to be available in its full force in the case of the opaque treatments discussed below. There should simply be no prohibition in principle to apply such automatisms without systematic human intervention.

**Définir une variable para\_noye = para\_noye + 1. Definition of transparent treatments.** It is still necessary to correctly define the field of these treatments which are claimed to be harmless, or at least free of the dangers which characterize opaque treatments. They could be defined as follows:

"An automated processing operation intended to produce an individual decision is said to be "transparent" when the nature and weighting of the input variables are fully fixed and described,

before its execution, in a framework decision, the content of which is accessible to the data subjects, accompanied, if necessary, by an explanation drafted in simple and clear language, and of which the individual decisions constitute the mere execution".

**Définir une variable para\_noye = para\_noye + 1. Right to an explanation and transparent treatments.** Let us admit that the automatism thus defined must escape the prohibition of principle which will strike the opaque treatments. In the current scheme, it is only when the processing operation was prohibited in principle, but falls within the scope of an exception that the data subject benefits from a set of guarantees. But there is nothing to prevent a complete change of approach, and to directly establish a right to "human intervention" in the case of transparent processing. But the nature of this intervention must be clearly understood. By hypothesis, there would be no sense here in appealing: by making the second decision, the human being would remain under the dependence of the scale, tariff or framework of which the first decision was only a circumstantial application. For the same reasons, offering the person concerned the opportunity to present observations in support of his or her appeal in order to ensure the adversarial nature of the procedure would be completely futile. One could certainly imagine that the controller would go over the form with the data subject to explain its meaning, but if the form is complex, he should have already published or made available such an explanation for the benefit of all. The value of individual human intervention is to verify and comment on the application of the scale or framework to the specific values in the individual's file. For example, when the tax authorities are asked about the amount of a housing tax, they remind the person concerned of the calculation formula and, for each of its elements, give the value that was used in the case in question, which allows the person concerned to see that the total does not contain any error<sup>39</sup>.

"The right to human intervention," in the context of "transparent" decision processing, could therefore be formulated as follows:

"The person concerned shall be informed, at his or her request, of the variables used to produce the individual decision against him or her, so that he or she can check that the framework decision has been applied to him or her without error.

**Définir une variable para\_noye = para\_noye + 1. Questioning the Framework Decision.** If it is the content of the Framework Decision itself that the decision addressees find problematic, the answer is not a matter for the GDPR. Depending on the area under consideration, administrative, civil, commercial, labor or consumer law must be invoked to have the allegedly unfair framework amended. The fact that a contractual or regulatory tariff is subject to automated individual enforcement does not change its nature.

**Définir une variable para\_noye = para\_noye + 1. Discrimination control.** Finally, what about the risk of discriminatory decision<sup>40</sup>? By hypothesis, it will be easy to check *ex ante* that such a

---

39 Experiment conducted by the author.

40 On this subject: Défenseur des droits (in partnership with the CNIL), Algorithmes: prévenir l'automatisation des discriminations, 2020 report; Cédric Villani (dir.), Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne, p. 147, which proposes the introduction of a *Discrimination Impact Assessment*: "... The aim would be to accompany the AIP with a similar mechanism for discrimination, a *discrimination impact assessment*, or DIA, to get AI designers to question the social consequences of the algorithms they produce"; G. Haas and S. Astier, "Les biais de l'intelligence artificielle: quels enjeux juridiques?", Dalloz IP/IT and Communication Repertory, July 2019.

treatment does not discriminate directly, since the criteria used are entirely public. However, it is still possible that it may produce indirect discriminatory effects<sup>41</sup>. It is then possible to draw inspiration from Article L. 1134-1 of the French Labour Code, by carrying out *ex post* verifications, by adjusting the burden of proof: if the data subject presents factual elements suggesting that the transparent automated processing is the source of indirect discrimination based on a prohibited criterion, it is up to the person responsible to prove that his or her decision is not based on such criteria<sup>42</sup>.

The list of criteria that should not be discriminated against should either be left to the states or recast at the European level. Currently, the GDPR focuses on the special categories of data in Article 9, which does not allow for example the classic discrimination based on gender<sup>43</sup>.

In the context of opaque automated processing, on the other hand, direct discrimination is quite possible. Generally speaking, they are more dangerous, which justifies a much more restrictive legal regime.

## B - The strict regime of opaque decisional treatments

**Définir une variable para\_noye = para\_noye + 1. Definition of opaque treatments.** What do we mean by "opaque treatments"? Their definition can be conceived very simply by reversing that of transparent treatments. While transparent treatments show the nature and the weighting of all the variables used, treatments are opaque when all or part of this information remains inaccessible to the recipients of the decisions. When does this happen?

First, when certain families of inductive automatisms are used. We have seen that there are learning techniques in which the exact criteria finally used for the classification are inaccessible to the user, and even to the designer.

To this intrinsic and irremediable opacity are added, then, situations in which the decision criteria are known, but voluntarily passed over in silence. Even if we use deductive automatism, which would be satisfied with deploying a set of rules conceived by the human mind, or inductive treatments capable of giving an intelligible account of the choices they make, the data controller or the software designer sometimes *chooses* obscurity. There are legitimate reasons to do so, for

---

41 Within the meaning of Article 1<sup>er</sup> of Law No. 2008-496 of May 27, 2008, containing various provisions adapting to Community law in the field of the fight against discrimination: "Indirect discrimination is any provision, criterion or practice that is neutral on its face, but which may, on one of the grounds mentioned in the first paragraph, place persons at a particular disadvantage in relation to other persons, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of attaining that aim are necessary and appropriate.

42 Art. L. 1134-1 al. 1 French Labour law Code : "When a dispute arises due to a failure to comply with the provisions of Chapter II, the applicant for a job, an internship or a period of training in a company or the employee must present evidence suggesting the existence of direct or indirect discrimination, as defined in Article 1 of Law No. 2008-496 of May 27, 2008, containing various provisions adapting to Community law in the field of the fight against discrimination.

43 This is very clear from the reading of Recital 71: "In order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use adequate mathematical or statistical procedures for profiling, apply appropriate technical and organizational measures to ensure, in particular, that factors leading to errors in personal data are corrected and that the risk of error is minimized and secure personal data in a way that takes into account the risks to the interests and rights of the data subject and prevents, inter alia, discriminatory effects on natural persons **based on racial or ethnic origin, political opinions, religion or belief, trade union membership, genetic or health status, or sexual orientation**, or which result in measures having such an effect" (emphasis added).

example to protect the details of an original decision-making process, the efficiency of which provides a competitive advantage, and in which time and resources have been invested. But obscurity can also be a means of concealing the unmentionable, such as the use of a prohibited discriminatory criterion. In all cases, opacity must be assumed, and will produce legal consequences.

This approach has the added advantage of being technology-neutral, as is the GDPR in general. It would be a mistake to tie oneself to a transitional state of the art, for example by targeting specific families of *machine learning* or "artificial intelligence", whose contours would be both contestable and evolving.

The definition could therefore be worded as follows: "An automated processing operation intended to produce an individual decision is said to be 'opaque' when the controller deprives the data subject of an exhaustive description of the rules applied to him or her, either by choice or as a consequence of the techniques employed by the processing operation".

**Définir une variable  $\text{para\_noye} = \text{para\_noye} + 1$ . Maintaining the prohibition in principle.**

Since opaque treatments have been defined, we can fully apply to them the prohibition in principle of being exclusively at the origin of the most important decisions - those producing legal effects or significantly affecting the person in a similar way - whereas open treatments were free to use.

However, as the first part of the study suggested, care must be taken to avoid presenting systems that are in fact in a position to decide on their own as mere "decision aids". To this end, the data controller will be required to describe, as part of his transparency obligations, "the specific contribution of the human decision-maker intervening after the automatism, in particular the methods and criteria on which he bases his assessment". If he fails to do so, the data controller will be presumed to base his decision exclusively on the automated processing. He will therefore be hit by the prohibition in principle.

**Définir une variable  $\text{para\_noye} = \text{para\_noye} + 1$ . Role of states in exceptions.** As the law currently stands, there are several exceptions to the prohibition principle. Should a change be proposed in this regard? Let us recall the decision rendered by the Constitutional Council, examining the law that adopted French domestic law to the GDPR. Recital 71 had attracted the attention of the doctrine, which decided: "(...) the controller must ensure that he has control over the algorithmic processing and its evolution in order to be able to explain, in detail and in an intelligible form, to the data subject the way in which the processing has been implemented with respect to him. As a result, algorithms that can revise the rules they apply without the control and validation of the controller cannot be used as the exclusive basis for an individual administrative decision"<sup>44</sup>. For the Council, some decisions are so high-stakes that they do not tolerate opacity. The fact that a "human intervention" may be requested at a later stage will not change anything: it is not acceptable that, in the first instance, the situation of the person concerned is determined by inaccessible mechanisms. It must be borne in mind that the guarantee of human intervention is not intended to be invoked systematically by *all the* recipients of individual decisions - otherwise the automatism would have been completely useless. Some fates would therefore have been sealed without knowing how or why.

---

44 Decision No. 2018-765 DC of June 12, 2018.



Of course, the reservation of interpretation formulated by the Council is compatible with current European law. Indeed, among the exceptions to the prohibition of principle, both the "data subject's explicit consent" and the necessity "for entering into, or performance of, a contract" are inapplicable to administrative matters<sup>45</sup>. The only remaining exception, the law of the Member State, is by definition open to a specific national approach. But this reasoning cannot be transposed to private law, for which the other two exceptions are available. Yet, is it not conceivable that in certain asymmetrical private law relationships - labour law, consumer law, insurance law - it should also be considered that the most high-stakes decisions are not susceptible to being delivered by a black box? For these cases, if one wants to avoid potentially difficult discussions on the applicability of exceptions - on the qualities of consent, on the absolute necessity of recourse to an automatism for the formation or performance of the contract - the simplest thing to do is to stick to the prohibition principle.

We have certainly regretted, in previous writings, that the GDPR leaves too much room for national interpretation. But these regrets concerned *pure* questions of personal data law. For example, it is absurd to be able to set the age of "informational majority" for children at between 13 and 16 at the discretion of each member state<sup>46</sup>. On the other hand, the issue of automated decisions is not purely a matter of personal data law, but is necessarily mixed with national sectoral law, depending on the area in which it occurs.

It therefore seems desirable to us that the exceptions of "necessity for the contract" and "consent" should only be open to the data controller in the absence of opposition from national law. The use of black boxes to bring to light decisions with legal consequences or significant impacts of another nature on individuals would not be a right for data controllers, but rather a tolerance, which could be refused.

**Définir une variable para\_noye = para\_noye + 1. Safeguards.** Let us assume, however, that one of the exceptions can be invoked by the data controller without being opposed by national law. For example, a company that has published a job advertisement receives so many applications that it is forced to carry out an automated "pre-sorting". What are the applicable guarantees?

**Définir une variable para\_noye = para\_noye + 1. Right to an individual motivation.** The first step here is again to explain the decision that has just been taken. In the case of transparent automated processing, the exercise was simple, since it was only a matter of explaining how a framework formula was applied to individual values. In the case of opaque processing, the need for clarification is much greater, but paradoxically more difficult to satisfy. If the treatment at work is "self-teaching" and its own logic is impenetrable, the exercise seems, at first sight, impossible. It is not necessarily so. It is quite conceivable, for example, that a sophisticated algorithmic CV selection process looks for several distinct qualities. Perhaps it evaluates in one way the adequacy of the candidate's initial education to the proposed position (by comparison with a database of keywords), in another way his professional experience (by weighting former employers differently

---

45 The G29 guidelines on consent rightly state that the structurally unbalanced relationship between the administration and the citizen is in principle incompatible with the expression of a serene consent or refusal: Guidelines 5/2020 on consent within the meaning of Regulation (EU) 2016/679, May 4, 2020, n° 16. For a commentary on this text: N. Metallinos, "G29 Profiling Guidelines. An attempt to clarify the special regime applicable to profiling and automated decision making," CCE, 2018, no. 2, comm. 14.

46 Art. 8.1 of the GDPR.

according to their notoriety or size), and by yet another process the quality of his written expression as it emerges from a cover letter (by relying on machine learning from letters labeled "good" or "bad"). In this example, perhaps it is possible to explain to the candidate that despite an initial education that was considered satisfactory, the work experience was considered insufficient overall. When the opaque treatment is deductive, it will be all the easier to subdivide the criteria he or she is instructed to apply into broad headings, in order to bring out a motivation.

We therefore consider, contrary to the current solution, that the data controller should be required to accompany the decision with an individual motivation<sup>47</sup>. This reasoning could itself be produced in a fully automated manner, without which the use of software would be useless. This statement of reasons will not be subject to a quality requirement - it is difficult to see how this can be judged - but one can count on a structural incentive for the data controller to ensure that it is the best possible: an unconvincing statement of reasons will be a spur to the data subject to seek "human intervention", i.e. to lodge an appeal.

To those who believe that a decision whose inner workings remain opaque is necessarily an inexplicable decision, we propose a new parallel with the human decision. A recruitment jury has as many "black boxes" as there are members. Everyone will appreciate the reading of a CV or the conduct of a job interview in the light of a long experience forged from a body of learning. It will sometimes be impossible to know *exactly* why a candidate seemed "clearer" or "more skilful", but they will be able to report on this general diagnosis.

If it is really not possible to bring out any individual motivation for the processing, then the person responsible will have to assume before the recipients of these decisions that he is unable to provide a beginning of explanation for his choices.

Individual motivation, even if automated, will usefully complement the current requirements for global information on the "underlying logic" of the treatment, and will make it much more concrete.

**Définir une variable para\_noye = para\_noye + 1. Human intervention.** If this individual motivation, probably automated, has not convinced, the "human intervention" phase must begin. Without upsetting the solutions currently in use, it is simply necessary to specify the details and the chronology, for example in the following way:

"The data subject shall have the right to lodge an appeal after having been informed of the decision and the accompanying statement of reasons. He or she is then invited by the controller to submit written observations in support of his or her case. A new reasoned decision is taken by a human being, which completely replaces the first one. The reasons for the human decision cannot be based on the results of the opaque automated processing.

The last sentence must be understood. Of course, it is not forbidden for a human being to reach the same conclusions as the machine, but he or she must do so in his or her own way. The diagnosis that the candidate's initial training was not sufficiently in line with the position offered was perhaps

---

47 On the question of whether the current texts lead to an individual or collective right to explanation, ex post or 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, Dec. 2018, 1. For a more collective conception of the right to intelligibility of processing: CNIL, Comment permettre à l'homme de garder la main ?... , report cited above, p. 53.

excellent. But this must be demonstrated directly from the CV. The "computer said so" may be a clue for the human investigator, but certainly not an argument.

Of course, if all recipients of automated decisions request human review, then recourse to the machine will be futile. This is already the case in positive law. Automated individual motivation is a powerful incentive for the controller to have even a rough idea of the reasons for the classification of cases.

**Définir une variable para\_noye = para\_noye + 1. Control of discriminations.** It remains to ensure that the automatism is not a source of prohibited discrimination. In the case of transparent processing, the discrimination, by hypothesis, could only be indirect. In the case of opaque processing, direct discrimination is possible, since the criteria on which the decision is based remain concealed. This is not a risk specific to computers: let us recall once again that a human jury is a set of black boxes, which can conceal their sexist or xenophobic motivations behind a façade. Here again, the burden of proof must be adjusted: if a data subject demonstrates that discrimination is likely, it is up to the data controller to prove that it did not occur - which is a new incentive to master at least the broad outlines of the operation of its processing. The discussion will be informed by the production of a report showing whether the presence of the disputed criterion has a statistical impact on the meaning of the decision taken. It is highly recommended that data controllers carry out regular *ex-post* checks of this kind on their own<sup>48</sup>. This would allow them to quickly identify that a processing operation is producing anomalous results, and then to investigate the causes of this behavior.

\*\*

Let us close this study with a concrete proposal for the rewriting of Article 22 of the GDPR, which takes advantage of all the previous observations and synthesizes them.

## CONCLUSION

### Proposed addition to Articles 13 and 14 of the GDPR

X. When the purpose of the processing is to assist in making a decision that has legal effects on the data subject or that significantly affects him or her in a similar manner, the data controller describes the specific contribution of the human decision-maker involved after automation, in particular the methods and criteria on which he or she bases his or her assessment.

### Proposal to rewrite Article 22 of the GDPR

1. This article applies to decisions that produce legal effects with respect to the data subject of the processing of personal data or that significantly affect him in a similar manner. Decisions are

---

48 The above-mentioned EDPB guidelines include as "good practice" (p. 36): "regular quality assurance checks of their systems to ensure that individuals are treated fairly and not discriminated against, whether on the basis of particular categories of personal data or otherwise (...)".

considered to be those processes whose purpose is to choose between several possible actions or abstentions with respect to the data subject.

Each Member State may draw up a list of criteria which may not be the subject of any direct or indirect discrimination in decision-making with regard to the persons concerned.

2. An automated processing operation intended to produce an individual decision is said to be "transparent" when the nature and weighting of the input variables are fully laid down and described, prior to its execution, in a framework decision, the content of which is accessible to the data subjects and of which the individual decisions constitute the mere execution. If the framework decision is not directly formulated in simple and clear language, it is accompanied by an explanation that has these qualities.

The person concerned shall be provided, at his or her request, with the state of the variables used to produce the individual decision against him or her, so that he or she can verify that the framework decision has been applied to him or her without error.

Individual or collective challenges to the content of the Framework Decision are not covered by this text.

If a data subject or a group of data subjects presents factual elements suggesting that the transparent automated processing operation is the cause of indirect discrimination based on a criterion referred to in 1. of this article, the controller shall have the burden of proving that its decision is not based on such a criterion, after producing a statistical report which makes it possible to verify whether there is a correlation between the presence of the disputed criterion and the meaning of the decision.

3. Automated processing intended to produce an individual decision is said to be "opaque" when the controller deprives the data subject of an exhaustive description of the rules applied to him or her, either by choice or as a result of the techniques employed by the processing.

4. It is prohibited to base a decision that produces legal effects for the data subject or significantly affects him/her in a similar way exclusively on opaque automated processing.

A controller who claims to use simple decision support systems shall nevertheless be deemed to base his decision exclusively on automated processing if, in fulfilling his transparency obligations under Articles 13 and 14 of this Regulation, he has not indicated what the human decision-maker's own contribution is after the automation, in particular the methods and criteria on which he bases his assessment.

5. By way of exception to paragraph 3, the decision may be authorized by Union law or the law of the Member State to which the controller is subject and which also provides for appropriate measures to safeguard the rights and freedoms and legitimate interests of the data subject.

6. By way of exception to paragraph 3, the decision shall be authorized in the following two cases, unless the law of the Member State to which the controller is subject precludes it. The decision :

(a) is necessary for the conclusion or performance of a contract between the data subject and a controller ;

or ;

(b) is based on the free and express consent of the person concerned.

7. In the cases referred to in paragraph 6, the person concerned shall benefit from at least the following guarantees:

(a) the decision is accompanied by an individual motivation, which may itself be produced exclusively by automated processing;

(b) after having taken cognizance of the decision and the accompanying statement of reasons, the data subject shall have the right to lodge an appeal. The data subject is then invited by the controller to submit written comments in support of his/her case. A new reasoned decision is taken by a human being, which completely replaces the first one. The reasons for the human decision cannot be based on the results of the opaque automated processing;

c) if a data subject or a group of data subjects presents factual elements suggesting that the opaque automated processing is the cause of direct or indirect discrimination based on a criterion referred to in 1. of this article, it shall be incumbent on the person responsible to prove that his or her decision is not based on such a criterion, after producing a statistical report that makes it possible to observe whether there is a correlation between the presence of the criterion at issue and the meaning of the decision.