Algorithms, AI systems and public services: what rights do users have?

CRITICAL CONSIDERATIONS AND RECOMMENDATIONS

Ensuring that the law forgets no one

Défenseur des droits
RÉPUBLIQUE FRANÇAISE

Algorithms, AI systems and public services:
what rights do users have?

Critical considerations and recommendations

OPENING REMARKS

As algorithms and artificial intelligence (AI) systems continue transforming the face of public services and public officials' work, the Defender of Rights believes, as the institution responsible for defending the rights of people using public services, that this subject warrants closer examination.

With algorithms and AI systems becoming such an integral part of our lives, there is an abundance of literature focusing on their technical and legal development, as well as the ethical, legal, environmental, social and commercial issues that they evoke.

In the same vein as the other reports that take a closer look at digitisation in the public sector[1], the Defender of Rights has begun contributing to what is a necessary public debate on these systems. Algorithmic systems may herald a source of progress, but they also pose major risks to rights and freedoms, as the Defender of Rights already emphasised in two reports, one in 2020 on the risk of automated discrimination caused by algorithms[2] and the other in 2021 on the urge to safeguard fundamental rights in the use of biometrics[3].

The determination to press ahead with rolling out algorithms and AI systems on an increasingly massive scale across public policies and public services is having a substantial effect on both public action and the rights of the users concerned. Growing numbers of individual administrative decisions are partially or fully automated and are based on the results produced by algorithms or AI systems.

The rate of complaints that the Defender of Rights has received on these particular issues is still relatively thin on the ground. Complainants tend to be more concerned about the actual outcome of the decisions concerning them, rather than the process that led to those decisions. This finding can be explained by the fact that the decision-making process, including its automated nature, mechanisms and even any biases, is not directly visible. However, some of the complaints received point to potentially systemic problems, especially with Affelnet, the online procedure for allocating college places.

In this respect, court rulings as well as scientific research and work on legal doctrine highlight persistent problems in the wake of the many initiatives spearheaded by France's data protection authority (CNIL)[4], the Conseil d'État[5] and other public[6] or private research or training institutions[7]. In particular, questions are raised about the substance of human involvement in cases where the individual administrative decision is said to be only "partially automated", as well as the nature of the explanations provided to the user concerned about the operation of the algorithm or AI system by means of which the decision was - wholly or otherwise - taken.

Based on this knowledge and the observations and proposals already made, the Defender of Rights has decided to dedicate a report to these issues. The aim is also to reaffirm the perception of a public service that should always retain its role of delivering support and providing a service to the public, by respecting the key principles of ensuring service continuity, equality of service for all and adaptability to users' needs. As such, the institution is promoting a genuine user-oriented "culture of rights" across all government agencies. This implies taking steps to ensure that the objectives assigned to public service managers and frontline staff alike are focused on respecting users' rights and consequently not allowing government services to direct all their efforts, particularly when it comes to deploying algorithmic systems, on the priority of reducing costs and personnel, and improving their profitability indicators.

The growing inroads that algorithmic tools are making into public action are being fuelled by the trend of digitising administrative procedures. As far as users' rights are concerned, examining the issues raised by algorithm-driven public action should not obscure the importance of the specific problems associated with the "mere" digitisation of public services. The Defender of Rights' 2023 annual activity report emphasises that access to rights could be undermined, especially by digitisation which excludes some 10 million people[8]. The report also advises that digitising administrative procedures should supplement, rather than replace, traditional counter services, paper mail and the telephone in order to guarantee fair access to public services[9]. Therefore, the institution will continue following up on the recommendations that it has issued in this respect[10].

If the legal framework governing the use of AI systems and algorithms changes, the following rules will apply, either cumulatively or otherwise:

- Where personal data are processed, the obligations and safeguards set out in the French Data Protection Act[11] (regularly amended, such as in 2018 to bring the act into line with the General Data Protection Regulation, known as the GDPR[12]), for which the CNIL is the competent authority

- The obligations laid down in the 2016 Law for a Digital Republic[13], which have since been enshrined in the Code of Relations between the Public and the Administration (CRPA), whenever a government agency uses an algorithm to make or help make an individual administrative decision, whether or not personal data are processed

- Specific obligations associated with certain sectors, such as access to higher education[14]

In addition, the obligations arising from the Artificial Intelligence Act adopted by the European Union in 2024 (EU AI Act for short)[15] to regulate the development, marketing and use of AI systems, will be phased in from 2025. In 2030, all the rules relating to the high-risk AI systems listed in Annex III of the EU AI Act[16] will apply, especially the AI systems used in the fields of education, employment, access to essential public services, law enforcement, migration, and the administration of justice.

Finally, in the summer of 2024, the Council of Europe adopted a framework convention on artificial intelligence and human rights, democracy and the rule of law[17], which the European Commission signed in September 2024 on behalf of the European Union.

This report, which is intended for the Government, Parliament, public officials (government agencies and local authorities), associations for access to rights, and actual users, does not aim to address all the issues raised by the deployment of AI systems and algorithms in public services, and all the associated risks - especially concerning discrimination - which will be covered in subsequent publications. It is specifically interested in those cases where decisions are taken with support from algorithmic systems and AI, and also the resulting problems of human intervention on the one hand and the transparency and explainability of these decisions on the other. This report is designed to contribute to the debates that should continue to be held with a view to respecting the rights that already apply, or those that should apply, for public service users.

CONTENTS

PART I

THE DEPLOYMENT OF ALGORITHMS AND ARTIFICIAL INTELLIGENCE (AI) SYSTEMS BY GOVERNMENT AGENCIES IN THEIR RELATIONS WITH THE PUBLIC

A few definitions are required before this report provides a brief overview of the increasingly widespread use of algorithms by the authorities ("Mainstream use of AI in public services") and, at the same time, the gradual move towards a fully automated process for taking individual administrative decisions.

1- PRELIMINARY DEFINITIONS

There is **no legal definition** of what an algorithm actually means, but based on the work by the CNIL and France's Commission on Access to Administrative Documents (Cada), **an algorithm** can be defined as *"the description of a finite and unambiguous sequence of steps (or instructions) for obtaining a result from a set of inputs. For example, using a spreadsheet to determine a person's entitlement to a benefit on the basis of predefined criteria constitutes an algorithm."*[18] It is worth noting that: *"Before an algorithm can be implemented by a computer, it must be expressed in machine language and transcribed into a program (a text comprising written commands, also known as "source code"). This program can then be executed as software or compiled in the form of an application. Software generally uses a large number of algorithms for entering data, calculating and displaying the results, communicating the results to other software programs, and so on."*[19] It can be compared to a cooking recipe. In other words, the system takes the inputs (ingredients) and then carries out various operations on those inputs (dosing, mixing, cooking, *etc.*) to produce a result (a dish).

Algorithms have been in use for decades, such as France's Directorate General of Public Finance (DGFiP) and its algorithm to calculate income tax[20].

The law and regulations lay down a number of obligations for organisations, and especially government services, that use algorithms for making individual decisions.

New challenges are emerging with the advent of the third industrial revolution and big data. *Big data* encompass all the data that people generate in their daily lives when using digital tools (messages, GPS signals, photos, transaction data, etc.). These data, which are created by and about individuals, include voluntary data, observed data (our digital footprints) and data inferred from voluntary and observed data. Data science is the discipline that studies information and the methods used to transform that information into "meaningful" resources.

AI systems and algorithms may or may not be based on personal data processing.

There are currently two main types of algorithm:

- **Closed algorithms,** which are classic tools. All the instructions have been thought out and written by a human being in source code. They combine only ad hoc data that have been determined in advance, i.e. data that the person responsible for the algorithm has previously identified and that the data subject has voluntarily provided on a one-off basis for a specific purpose (such as the Parcoursup university admissions portal, where the list of personal data processed by the algorithm is set out in a regulation).

- **Machine learning algorithms**, which operate via a process whereby relevant information is drawn from a set of training data, often with the aim of predicting or estimating an outcome. The human only tells the algorithm what result should be achieved, based on the training data that can be provided. The program then automatically generates the operations needed to attain the required result.[21]

- Over the last 15 or so years, the **deep learning** method has been developing. This automatic learning process uses artificial neural networks featuring several layers of neurons. These algorithms contain vast numbers of parameters and require tremendous amounts of data to be trained[22], i.e. to build a model from the data.

Learning algorithms form the backbone of what are known as AI systems. According to the AI Regulation adopted by the European Union in 2024, an AI system is *"a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."*[23]

These algorithms pose a number of challenges:

- The use of algorithms, whether simple or learning, raises transparency issues, since the people using those algorithms and also the people concerned by their use must be able to gain a certain level of insight into the role that these systems played in the decision taken and their internal logic.

- Since learning systems are defined by their ability to develop their own rules for achieving the required result, they present a specific challenge when it comes to opacity. Even the people designing these algorithms *"lose the ability to understand the logic behind the results produced."*[24]

These issues will be discussed in greater detail in the final section of this report.

Closed algorithms and AI systems, along with examples in various sectors of public action, are examined in this report, since they are increasingly used to make individual administrative decisions.

2· MAINSTREAM USE OF AI IN PUBLIC SERVICES

In the same spirit as the trend of **digitising administrative procedures**, which can be defined as the growing use of online procedures in public services, the authorities have accelerated their plans to roll out algorithms and AI systems in government services. The rise of big data, in particular, paves the way for this kind of deployment.

Many government services are already based on algorithms to such an extent that some authors have spoken about the **algorithmization of government services[25]** or **algorithmic public action[26]**. They refer to the authorities' growing reliance on algorithms or AI systems for automating, standardising, improving and/or speeding up certain administrative procedures or certain parts of those procedures.

INSIGHT

DEDICATED STRUCTURES POWERING THE GOVERNMENT'S DIGITAL TRANSFORMATION

Over the last few years, various government agencies with ever-changing responsibilities have been guiding efforts to roll out algorithmic systems and digital technology across public services. Since 2019, such interministerial action has been led by two main directorates:

· The **Interministerial Directorate for Public Transformation (DITP)**, which leads and coordinates the public transformation programme and oversees the implementation of the government's priority policies. It supports ministries with their transformation projects, and designs and implements key interministerial projects for transforming the State.

· The **Interministerial Directorate for Digital Technology (Dinum)**, which is responsible for developing and guiding the State's digital strategy. One of its objectives involves developing the effective use of data "to make the State more effective in its actions and simpler for citizens, businesses and public officials", especially through the development of algorithmic processing[27].

> Within the Dinum, the **Etalab** department created in 2011 is tasked with coordinating the design and implementation of the State's data strategy with three focus areas:

· Open data and data sharing: Etalab coordinates the actions of the various government agencies and provides them with support in establishing easier methods for disseminating and reusing their public information.

· Data mining, artificial intelligence and public algorithms: Etalab coordinates and promotes the action of the State and the bodies under its authority in terms of mapping, governing, producing, circulating, using and opening up data, particularly source code.

· Innovation and open data in public action: Etalab is helping open up government services and public action to society by implementing the principles of "open government" (transparent public action, consultation with civil society, citizen participation, and open innovation).

Etalab supports government agencies in opening up and making responsible use of public algorithms. For example, it published a Guide to Public Algorithms[28] in 2019 for all government agencies and public service organisations that design, develop and operate algorithmic processing operations, to offer guidance on applying the legal framework relating to public algorithms, especially with regard to the requirement for transparency.

Etalab has categorised the different types of algorithm use identified in public services:

USES	EXAMPLES
Allocate rights, calculate amounts according to predefined rules	Calculate taxes and levies, grant entitlement to social benefits, etc.
Match supply and demand	Manage staff mobility (HR), allocate places in higher education (Parcoursup), allocate places in nurseries, etc.
Predict a situation or risk by analysing data	Predict the risk of a defaulting business (weak signals), target efforts to tackle tax fraud, etc.
Help users make decisions	Help jobseekers target their unsolicited applications ("La Bonne Boîte" website), simulate the cost of hiring, etc.

Government agencies, local authorities, the justice system and public services in general are **already making increased use of AI systems and algorithms**, either on their own or with research institutes or private service providers, sometimes on a trial basis, to analyse large volumes of data for such purposes as:

- Anonymising court decisions[29]

- Identifying events in the public space for "security" reasons using augmented CCTV systems [30]

- Improving medical diagnoses and preventing illnesses[31]

- Personalising learning[32]

- Tackling food waste in restaurants within public-access buildings[33]

- Guiding the deployment of public policies, and simulating the potential economic and social consequences of a new legislative measure[34]

- Analysing local needs (managing water resources and preventing/detecting leaks, or implementing and improving smart lighting systems)

Developments involving the use of generative AI systems are currently grabbing the headlines. According to the CNIL: "*Generative artificial intelligence is a system capable of creating text, images or other content (music, video, voice, etc.) from a human user's instruction. These systems can produce new content from training data.*"[35]

While ChatGPT has now become a household name, several generative AI systems have been developed in the field of public services. For example, a generative AI system has been designed to summarise the amendments tabled to bills or proposed laws (known as "LLaMendement"). As far as relations between users and the authorities are concerned, "Albert" (another example of a generative AI system) has been in development since June 2023.

"ALBERT", THE GENERATIVE AI SYSTEM[36]

> Albert is the name that has been given to a **generative** AI system developed by the State to "**fast-track administrative formalities** and provide **public service** users with reliable, clear and efficient answers."
>
> Dinum is developing this chatbot using existing resources (especially the fact sheets on the service-public.fr portal). At this stage, Albert is reserved for public officials to help them deliver a better response to users' enquiries, search for information about regulations and even produce summaries. It is starting to be deployed across the France Services network of walk-in information centres.
>
> There are already plans to use the system for other applications:
>
> - **Automatically subtitling lectures**
> - **Transcribing court hearings**
> - **Drafting complaints or producing medical reports**

There are two reasons why special care must be taken with the government's use of these procedures[37]:

- **Algorithms are used to enforce regulations** (laws, statutes, etc.)
- Unlike a private service, which people can opt out of using (to a certain extent), **public service users have no such choice**.

Having said that, although the deployment of "algorithmic" public action may raise a number of questions, there may be some advantages when government agencies use automated systems in their relations with users[38]:

- It can lead to more effective and better informed decisions by processing a wealth of data.
- It may involve establishing and formalising criteria that can be used to identify and clarify the logic behind the decisions taken, where those same decisions may previously have been taken in an opaque manner (such as allocating nursery places).
- As such, it can help reduce the arbitrary nature of certain decisions and promote equal treatment, since human agents are not always neutral and their decisions can prove to be discriminatory.

For these advantages to become a reality, AI systems and algorithms must be implemented and used in a responsible manner and in accordance with the law, otherwise critical problems could arise.

3· USING FULLY AUTOMATED PROCESSES FOR TAKING INDIVIDUAL ADMINISTRATIVE DECISIONS

Ever since the French Data Protection Act became effective on 6 January 1978, **fully automated individual decision-making has been prohibited within public services**. This intention by the legislature was already apparent when it was entrusted by the Senate's Constitutional Law Committee to produce a report on the proposed data protection act. In view of the increasing use of computers, the report stressed the need to *"ensure that this method of judgment in no way replaces the traditional methods or introduces automated processes where nuance, not to say delicacy, is often required."*[39]

Despite the various amendments made to this act, this principle of prohibiting automated decision-making still explicitly applies to **court decisions** involving "*an assessment of a person's behaviour*."[40]

The same cannot be said of administrative procedures.

While it was accepted that government agencies already relied on algorithms to help make decisions (such as for calculating entitlement to benefits, assigning human resources, or allocating places for pupils and students), **the Government's intention in 2018 was to *"open up more widely the possibility for public services to use automated decisions** (taken on the basis of*

an algorithm)."[41] This was achieved by enacting the country's law on personal data protection[42].

In the same way, the impact study into the proposed data protection law emphasised that *"human intervention does not always appear to be necessary to safeguard the rights of data subjects and public service users, such as when calculating personal income tax. Maintaining such an outright ban thwarts efforts to respond to changes in government agencies' activities, which are making increasing use of algorithmic processing operations, especially mass decisions for which the regulations provide a precise framework and whose rapid delivery could enable those agencies to render a proper public service."*[43] The study highlighted two examples in the field of public finance, namely the calculation of levies and tax on the one hand, and the fight against fraud on the other. The underlying idea was that this type of decision-making, subject to certain conditions, was key to delivering efficient public services.

Therefore, the principle of prohibition has been accompanied by **exceptions that open up a wide range of possibilities** for using **fully automated individual administrative decisions**. As far as the authorities' actions are concerned under current law:

- There is still a principle of prohibiting fully automated decisions that *"produce legal effects for individuals"* or which *"significantly affect them."*[44] This particular area is both vast and potentially complex to identify.

- In the wake of the GDPR[45] on a European Union level, the legislator decided[46] that **individual administrative decisions meeting certain conditions could be taken *"based solely on the automated processing of personal data."***

Note that the focus here is on individual administrative decisions that do not fall within the scope of criminal law within the meaning of Directive 2016/680 of 27 April 2016, known as the "Law Enforcement" Directive[47], which provides for an exception to the GDPR.

The conditions governing fully automated individual administrative decisions, which constitute **safeguards for public service users**, are as follows:

1. **Inform the person concerned** that the decision has been taken on the basis of an algorithm and provide the main elements for understanding the algorithm.

2. **Ensure that the algorithmic processing operation is overseen** by the public body using it.

3. Allow the person concerned by the decision to **obtain human intervention in the event of an appeal** against the fully automated decision.

4. Prevent fully automated decisions from being taken on the **basis of "sensitive data"** within the meaning of the French Data Protection Act.

The Constitutional Council approved these provisions by clarifying the conditions laid down by the legislator, particularly by advising that *"algorithms may not be used, as the sole basis for an individual administrative decision, where they are capable of revising the applicable rules themselves, without oversight and validation by the data controller."*[48] Consequently, it is on the basis of the safeguards provided that the exception allowing for administrative action has been deemed to comply with the requirements underlying our legal framework.

More than five years after the law was enacted, the time has come to review the overall balance of the system and the implementation of these safeguards.

SENSITIVE PERSONAL DATA AND THE RISK OF AUTOMATED DISCRIMINATION

The risk of automated discrimination caused by the use of algorithmic or AI systems has been widely demonstrated[49]. There are very real examples where this risk has unfortunately come true, including in the public sector.

> These cases include the algorithm developed in Austria by the national employment agency "AMS" in 2019 with the aim assigning an "employability score" to the unemployed. Statistical regression analyses were carried out to determine the factors that are most likely to predict an individual's chances of finding a job and the type of support that should be offered in light of the agency's limited resources.

According to the survey carried out by the AlgorithmWatch[50] association, the system deployed automatically assigned a score to each jobseeker based on several attributes. Jobseekers were divided into three groups according to their score:

- Group A for people who do not need any help finding a new job
- Group B for people who might benefit from retraining and support
- Group C for people deemed unemployable, who will receive less help from AMS and may be referred to other institutions

Once the system's results were made public, the system received a backlash for its discriminatory bias and its failure to uphold the right to non-discrimination. For example, unemployed women were given a negative weighting, as were people with disabilities and people over the age of 30. Women with children were also negatively scored, which was not the case for men with children. In so doing, the algorithm endorsed the systemic discrimination and inequalities in access to employment already seen in the market. The system, which was initially declared illegal by the Austrian data protection authority, was accepted on appeal.

However, the AMS algorithm is no longer used following the controversy that it caused.

Among the safeguards that were adopted in France when authorising fully automated individual administrative decisions, **a ban on basing such decisions on the processing of so-called "sensitive" personal data was laid down as a means of combating potential discrimination[51]**.

Article 9 of the GDPR defines such processing as follows: *"processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the 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."*

While this definition seems clear, its recent interpretation by French and European courts raises a number of questions. In particular, it appears that:

- In a number of recent decisions, the Court of Justice of the European Union (CJEU) has tended to take the view that this category should be interpreted broadly, including when it is based on an indirect approach[52].
- The status of personal data that reveal the presence of a disability[53], and which are therefore potentially comparable to personal data concerning health, may be transformed.

Extensive interpretations of the concept of sensitive data, which is primarily the responsibility of the CNIL under the authority of the Conseil d'État, could restrict government agencies' ability to use a fully automated decision-making process.

In addition, the GDPR's definition of sensitive data does not exactly overlap with the characteristics protected by non-discrimination law[54]. Family status (particularly single parents), economic vulnerability, gender and even place of residence are protected under non-discrimination law, but are not considered to be sensitive data within the meaning of the applicable data protection regulation.

Finally, it should be remembered that learning algorithms can incorporate indicators and equivalent variables (known as proxies) during their development. These proxies can be used as a substitute for a discrimination criterion. For instance, an indicator for people who have bought tampons can also be used to measure women (in a certain age bracket) and target them where appropriate. Such "redundant encoding", i.e. belonging to a protected category encoded in other data, is not necessarily intentional but results from the learning programme.

INSIGHT

THE CONCEPT OF "DECISION"

The need for a broad interpretation of the concept of "decision".

In a recent judgment[55], the CJEU examined the way in which a German company assigned a solvency score to a natural person, whose score was then transmitted to a bank. The Court considered that in the case where the probability value established by the company and communicated to a bank plays a determining role in the granting of credit, the establishment of that value must be qualified in itself as a decision producing, vis-à-vis a data subject, *"legal effects concerning him or her, or similarly significantly affecting him or her."*

The Court emphasised that this interpretation, and especially the broad scope of the concept of "decision", reinforces the effective protection intended by the GDPR. On the other hand, there would be a lacuna in legal protection if a restrictive interpretation was retained, according to which the establishment of the probability value must only be considered as a preparatory act and only the act adopted by the third party can, where appropriate, be classified as a "decision". In that situation, the establishment of a probability value would escape the specific requirements of the GDPR, which prohibits, as a matter of principle, fully automated decisions when this decision has legal or significant effects for the data subject, although exceptions are permitted. This procedure is based on automated processing of personal data and produces effects significantly affecting the data subject to the extent that the action of the third party to whom that probability value is transmitted draws strongly on it.

Based on this ruling, it will be necessary to determine whether the scope of individual administrative decisions, as stipulated in Article 47(2) of the French Data Protection Act, needs to be clarified.

The question already arises as to the "risk scores" that family allowance funds assign to their beneficiaries[56], with regard to the impact that this score has on the implementation of checks and inspections to combat error and fraud. The distinction currently made in applicable law focuses on "the decision" in a way that may seem inadequate, since it does not necessarily cover the actual process that led to the decision. The decision taken by the authorities may be the final step in a process involving many different stages. What happens if one of these stages, which is fully automated, plays a decisive role in the decision?

When ruling on the provisions of the 2018 law that ushered in fully automated individual administrative decisions, the Constitutional Council considered that government agencies were authorised to *"carry out an individual assessment into the situation of the person concerned, solely by means of an algorithm, according to the rules and criteria previously defined by the controller."*[57] This is the very issue at stake when analysing the need for human intervention.

PART II

OVERSIGHT AND HUMAN INTERVENTION: SAFEGUARDS, BUT WHAT SPECIFIC MEASURES?

As the law currently stands, one of the major issues when it comes to individual administrative decisions is establishing whether they are fully or only partially automated, which raises questions about the substance of human intervention.

This issue needs to be placed in a broader context, i.e. the extent of human control over automated systems that are used to make individual decisions and also guide public policy (such as for anticipating the spread of an epidemic and identify the most affected areas).

1· DIFFERENT LEVELS OF HUMAN INTERVENTION

The concept of "human supervision" refers to the need for humans to "stay in control" of the systems used[58] so that they do not operate in complete independence. It also refers to the fact that the responsible party - government agencies in this case - cannot invoke the involvement of the algorithm or AI system used to take an individual administrative decision (whether completely or partially automated) if it were challenged as part of a dispute.

Current regulations use a variety of ways to define this supervisory role. They describe different concepts and specify different obligations. For example:

- The concept of "control" by the controller is set out in Article 47 of the Data Protection Act of 6 January 1978.

- The concept of "human intervention" during automated individual decision-making or profiling is enshrined in Article 22 of the GDPR.

- The concept of "human oversight" has been introduced in Article 14 of the EU Regulation on AI.

These terms are used in this report as substitutes to refer to the human intervention or review that can and must be exercised at the different stages in the decision-making process.

Human control over these systems mainly occurs at two levels:

- "System" control, which concerns the operation of the system as a whole.

- "Individual" control, which concerns an individual decision.

"SYSTEM" CONTROLS

A distinction can be made between three separate control phases, namely during design (phase 1), during operation (phase 2) and during subsequent testing and audits (phase 3). These are three types of control that take place during phases which, in theory, are separate but may overlap (audits may take place during operation).

Human intervention from the outset when the rules are established by the algorithm or AI system (phase 1).

When the algorithm is implemented by a public authority, the choices made must comply with the regulations that apply in the area concerned, in addition to the legal framework that generally governs algorithms. For example, France's Biomedicine Agency uses an algorithmic system to determine the order in which heart transplants are proposed, known as the "heart score".

In practice, whenever a heart transplant becomes available, the system calculates a score for all the people on the transplant waiting list. In accordance with sector-specific legislation, the rules for this algorithm must prioritise the under-18s *"pursuant to the procedures defined by the Biomedicine Agency and taking into account the degree of urgency for a transplant."*[59]

A public service deploying an algorithm or AI system is responsible and accountable for its operating rules and criteria. In addition, the rules that define the main algorithmic processes that the public service uses to perform its missions, where these processes form the basis for individual decisions (whether in part or in full), must be published[60].

EXAMPLE

THE LOCAL ALGORITHM DEVELOPED AND USED BY SCIENCES PO BORDEAUX AS PART OF THE PARCOURSUP PROCEDURE

The Parcoursup web portal is the main procedure that students must use to secure a place on an undergraduate course at a higher education institution. It will be presented in detail below. Broadly speaking, the procedure involves three separate phases:

- Applicants register their choices on the national platform managed by the Ministry of Higher Education and Research.

- Higher education institutions classify the applications received over the national platform, often using a local algorithm to evaluate those applications (the tool proposed by the Ministry, or another solution), before sending back a ranking list.

- The algorithm on the national platform matches each applicant's choices with the application ranking lists provided by the higher education institutions, and then sends the proposals to each applicant.

Sciences Po Bordeaux states that it uses a different algorithm to the decision support tool provided by the Ministry for selecting applications[61]. As such, it cannot select any personal data other than those provided for by law[62]; however, it can use any criteria that it deems relevant.

In this case, the decision was taken to not rely (to a certain extent) on the student's grades alone (the criterion used for the decision support tool proposed by the Ministry), but on the difference between the student's grades and the class average. This solution, which enables Sciences Po Bordeaux to assess students' levels in greater detail, is therefore detrimental to students in colleges that tend to inflate grades and to those in classes featuring students of the same ability level, which is reflected in a narrow range of grades within the class[63]. This algorithm for selecting applications is also configured to favour students with grants (over and above the

quota required by law) and those from colleges participating in the "Cordées de la Réussite" scheme aimed at promoting greater access to higher education.

The algorithm's designer, Professor Vincent Tiberj, claims that it is "a sociological choice" and explains that the consequence of the choices made by this "in-house" algorithm is that it "favours those colleges that accept diversity within their classes".

> The results are striking. In 2023, the student profiles who were considered eligible using the "in-house" algorithm and the student profiles considered eligible using the decision support tool proposed by the Ministry varied significantly:

- Pupils on a "Cordée de la Réussite" scheme went from 19% in the first group to 3% in the second.

- The number of students with grants went from 12% to 5%.

- Students from overseas France went from 8.5% to 3.5%.

This case illustrates the fact that this algorithm, as with others, harbours a political choice (and not just a technical choice) that is made by humans, decisions that are taken, and even effects that are anticipated.

The way in which these algorithms are configured must comply with applicable regulations (such as respecting the list of personal data processed) and reflect the political objectives defined by the authorities concerned. Since there is a risk of a discrepancy between the legal and administrative framework and the code, which are not based on the same language[64], this implies developing a shared understanding of these tools and especially possessing the necessary technical skills (knowing how to code) and time, which government agencies cannot necessarily provide. This is especially true of higher education institutions, the vast majority of which are more than happy to use the decision support tool proposed by the Ministry for the Parcoursup admissions procedure.

In machine learning systems, the algorithm itself defines the rules. However, humans can get involved at the following stages (non-exhaustive list):

- Defining the objective to be achieved

- Selecting the data that will be used to train the system ("training data" or "training set")

- Labelling or annotating the data (this stage can also be automated and involves describing the data)

- Choosing the criteria used to assess the system's results (also known as "metrics"), such as the rate of false positives[65] and false negatives[66]

By way of example, in the field of waste management, local authorities can design AI systems to measure bin fill rates and adapt refuse collection schedules accordingly.

Human intervention in the sense of supervising the system during its operation and during subsequent testing and audits (phases 2 and 3).

The aim during these phases is to:

- Monitor how the system is performing and stop it at any time if a fault occurs

- Detect errors and biases, in addition to the debugging work that was already carried out during the design phase

These steps are particularly important when updating machine learning algorithms, since they have the special ability to revise their own rules and therefore evolve. Consequently, they contain the increased risk of an error on the one hand and discrimination on the other, due to the correlations that they create.

That explains why the EU AI Act has created obligations in this area.

PROVISIONS OF THE EU AI ACT FOR "HIGH-RISK" AI SYSTEMS

The EU AI Act subjects AI systems that are considered to be high-risk[67] (listed in Annex III of the Act) to the principle of "*human oversight*". In the words of the Act, the aim is to minimise the risks to health, safety or fundamental rights of an AI system deployed by the user[68] in accordance with its intended purpose - i.e. as specified by the provider - or under conditions of "reasonably foreseeable" misuse.

Some of the AI systems identified by the EU AI Act as "high-risk systems" concern the operation of public services, since they include:

- *In the area of access to and enjoyment of essential private services and essential public services and benefits, AI systems intended to be used:*

> By public authorities (or in their name) to evaluate the eligibility of natural persons for essential public assistance benefits and services, including healthcare services, as well as to grant, reduce, revoke, or reclaim such benefits and services.

> For risk assessment and pricing in relation to natural persons in the case of life and health insurance.

> To evaluate and classify emergency calls by natural persons or to be used to dispatch, or to establish priority in the dispatching of, emergency first response services, including by police, firefighters and medical aid, as well as of emergency healthcare patient triage systems.

- *In the area of education and vocational training, AI systems intended to be used:*

> To determine access or admission or to assign natural persons to educational and vocational training institutions.

> To evaluate learning outcomes, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions.

> For the purpose of assessing the appropriate level of education that an individual will receive or will be able to access, in the context of or within educational and vocational training institutions.

> For monitoring and detecting prohibited behaviour of students during tests.

Specifically, Article 14 of the Act relating to human oversight stipulates that high-risk AI systems must be designed and developed in such a way, *"including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use."* Article 14 also specifies that oversight measures must be commensurate with the risks and shall be ensured through:

- Measures identified and built, when technically feasible, into the high-risk AI system by the provider before it is put into service.

- Measures identified by the provider before putting the high-risk AI system into service and that are appropriate to be implemented by the deployer.

According to the Act, what that specifically means is that the natural persons responsible for carrying out human oversight on behalf of the organisation deploying the AI system must be enabled to:

- Properly understand the relevant capacities and limitations of the high-risk AI system and be able to duly monitor its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance.

- Remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons.

- Correctly interpret the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available.

- Decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the high-risk AI system.

- Intervene in the operation of the high-risk AI system or interrupt the system through a stop button or a similar procedure.

At the time of press, the harmonised standards that will be specifically and precisely used to support the requirements laid down in the AI Act, and especially the obligations set out in Article 14, are currently under development. The JTC-21[69] joint technical committee set up at the request of the European Commission is the body responsible for this process. Broadly speaking, AI systems that comply with these future standards will be presumed to comply with the obligations set out in the AI Act.

While the Defender of Rights welcomes the fact that the requirement for human oversight of the system during its use is defined in Article 14 of the AI Act, a close watch will need to be kept on the content of the future harmonised standard, which should specify the requirements of the Act.

In addition, Article 14 of the EU AI Act only covers AI systems that are considered to be high-risk, while those considered to be low-risk and minimum-risk are not concerned. It is worth keeping in mind that the Act provides for a filter system[70] that can downgrade an AI system that would otherwise be in the high-risk category based on the area in which it is used. A system can be removed from the high-risk category where any of the following (alternative) conditions is fulfilled:

1. The AI system is intended to perform a narrow procedural task.

2. The AI system is intended to improve the result of a previously completed human activity.

3. The AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review.

4. The AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III.

By way of exception, an AI system will always be considered to be high-risk where the AI system performs profiling[71] of natural persons.

On this particular point, the EU AI Act refers to the GDPR, which provides the following definition (Article 4(4)): "*profiling means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements." As the CNIL points out: "Profiling processing involves building an individualised profile of a specific person and aims to evaluate certain personal aspects for the purpose of judging or reaching conclusions about that person. Special care must be taken with any profiling processing operations since, by their very nature, they raise significant risks for the rights and freedoms of natural persons." The CNIL gives the following example: "Profiling underlies the insurance sector, particularly when it involves evaluating the characteristics of the insurance risk with the aim of determining the frequency, average cost, maximum cost of a potential claim, pricing and the insurability of the risk. Therefore, it is only a means and not an end.*"[72]

In this respect, the Defender of Rights is worried that the AI systems deployed by government agencies, local authorities and other public service organisations in their relations with users (both generally for guiding public policy, and individually for decision-making purposes) could be removed from the "high-risk" category as a result of this filter system. This is due to the risks identified and the safeguards that the regulation provides for this category.

The EU AI Act also sets out further obligations for high-risk AI systems, including:

- A risk management system (Article 9), relating in particular to the identification and analysis of the risks (known and reasonably foreseeable, and also likely to emerge) that the AI system can pose to, inter alia, fundamental rights when the system is used in accordance with its intended purpose, and the adoption of appropriate and targeted measures to address the risks identified.

- Data governance and management practices appropriate for the intended purpose of the AI system for training, validation and testing data sets, where the system makes use of techniques involving the training of AI models with data (Article 10).

- Maintaining technical documentation that has been drawn up in such a way as to demonstrate that the AI system complies with the requirements set out (Article 11).

- AI systems are designed and developed in such a way that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately, particularly by means of instructions for use (Article 13).

- An appropriate level of accuracy, robustness, and cybersecurity, and consistent performance in those respects throughout the lifecycle of these systems (Article 15).

INDIVIDUAL "CONTROLS"

Human intervention when individual administrative decision-making is based on the results of algorithms or AI systems.

Applicable law makes a distinction between two types of individual administrative decision, depending on the degree of automation or automaticity[73]:

- On the one hand, there are **partially automated individual administrative decisions**, i.e. decisions where an algorithm or AI system was used at a given moment and for a specific purpose, but where the result of those systems was only one element in the decision-making process, which took other aspects into account. Human intervention is especially required when the personal data being processed contain sensitive data according to the GDPR, which explains why it is so important to be clear about what types of data are classed or not classed as sensitive data.

  > Example: determining the amount of the disability compensation benefit managed by a given local authority[74].

- On the other hand, there are **fully automated individual administrative decisions**, i.e. decisions made by the authorities without taking into consideration any elements other than the sole result of the algorithm or AI system, which alone constitutes the decision. In other words, there is no human intervention during the individual decision-making process. However, from a legal perspective, the decision remains that of the government agency or local authority.

  > Example: calculating the amount of income tax that an individual is required to pay.

The criterion for separating these two decision-making categories is **human intervention. To qualify as such**, it must be consistent and have a real influence on the result.

Human intervention is compulsory in case of an administrative appeal against the individual decision.

In the event of an administrative appeal against an individual decision (in pursuance of Articles L. 410-1 et seq. of the CRPA), human intervention is required[75]. This appeal requires special attention from the authorities.

Therefore, this type of intervention does not concern "mass" processing operations and, with regard to the safeguards already provided for in Part I of Book IV of the CRPA Code, does not require any particular observations at this stage.

2· THE SUBSTANCE OF HUMAN INTERVENTION IN INDIVIDUAL DECISION-MAKING - RECOMMENDATIONS FOR GUARANTEEING INVOLVEMENT

To qualify individual administrative decision-making as only partially automated, human intervention cannot be limited to a *"token gesture"*[76]. In particular, such intervention may take the form of requiring an additional step, i.e. the official performs a positive, specific and meaningful action based on or alongside the result generated by the algorithm. For example, the official may be given the same task as the algorithm. If the results are different (e.g. calculation or identification), the official must then decide which of the two results to use.

Therefore, if an official consistently uses the results generated by the algorithm or AI system, such as by hitting a "confirm" key, without any intellectual checks or verifications, there cannot be considered to be any human intervention, and the decision-making process will not qualify as partially automated. Nearly 10 years ago, the Conseil d'État was already emphasising the need to avoid a situation where *"systems presented as offering decision support are in reality almost always followed and dictate the decision, and any element of human involvement is only apparent"*[77]. Consequently, significant leeway is required on the part of public officials to consider that automated processing only plays a partial role when reaching a decision.

Examining the CNIL's deliberations reveals elements of human intervention that relate to the existence of additional stages in the decision-making process. For example:

- *"Individual examination, especially to eliminate cases of mistaken identity, after collecting additional information where necessary (...)"*[78].

- Or the fact that some remote monitoring systems for online exams have the *"sole role (...) of drawing the invigilator's attention to a potentially abnormal situation,"* where *"human checks"* must always be performed before confirming any suspicions of cheating[79].

> In 2020, the Constitutional Council examined a set of provisions allowing the authorities to use computerised and automated processes to collect and exploit content available on the Internet for the purpose of combatting tax fraud. The Constitutional Council considered that there was no fully automated decision-making process involved and approved the new control system for collecting tax. According to the Constitutional Council, the *"corroboration"* and *"enrichment"* operations carried out by the competent authorities characterise *"an individual assessment of the person's situation by the authorities, which cannot then base its decision exclusively on the results of the automated processing operations."*[80]

It is clear from the above that two characteristics must always be present, i.e. the person who intervenes must have the authority to do so (which could be defined as the fact that the internal organisation recognises that this person has the capacity to act) and the competence (which could be defined as the fact that the person has the intellectual and practical means to carry out the review).[81]

Some authors have pointed out[82] that having the competence and authority to intervene in the result produced by the algorithm or AI system does not actually mean that there is human involvement in all cases.

In addition, according to legal expert Winston Maxwell, two other conditions must be satisfied for a human review to be considered effective[83]:

- Have knowledge of the operational characteristics and limitations of the algorithm, i.e. know which rules were used to design the algorithm and at what point in time, etc.

- Engage in a deliberative thought process that takes into account the context of the decision (think critically about the way in which the system works and produces its results), and think about whether this is an "atypical" case in light of the inputs used to train the algorithm.

Otherwise, human intervention may be provided for only on a purely formal basis. Entrusting individuals with the responsibility to supervise the operation of these systems runs the risk of *"turning humans placed in the loop into "moral crumple zones", largely totemic humans whose central role becomes soaking up fault, although they have only partial control of the system, and who are vulnerable to being scapegoated by tech developers and organisations seeking to avoid responsibility."*[84] Therefore, the objectives and resources made available for human review must be determined.

Generally speaking, two types of elements, which are now widely documented, lead to questions about the substance of human intervention in many so-called partially automated decisions.

On the one hand, the sheer volume of decisions in some cases raises suspicions about whether humans are truly involved in individual administrative decision-making processes.

The substance of such involvement is even more questionable when considering that algorithms and other AI systems are often implemented for the specific purpose of accelerating and, from the deployer's point of view, simplifying the decision-making process and preventing employees from having to carry out repetitive and time-consuming tasks, or even for downsizing the workforce.

THE CASE OF PARCOURSUP - LOCAL ALGORITHMS USED BY HIGHER EDUCATION INSTITUTIONS TO RANK STUDENT APPLICATIONS

As mentioned earlier, the Parcoursup admissions procedure comprises several stages. An essential stage is where the higher education institutions and more specifically the admissions board (CEV)[85] ranks the student applications received over the national platform.

Institutions can use an algorithm, whether the decision support tool proposed by the Ministry of Education or their own proprietary algorithm. As its name suggests, the Ministry's algorithmic tool should only help in the decision-making process, since the ranking decisions taken on the basis of this algorithm are supposed to be decisions that are not fully automated[86]. This was also the ruling that the Constitutional Council reached after the matter had been referred for its review. On the one hand, it held that the use of algorithmic processing by universities to rank student applications is merely an option, and on the other hand it considered that when universities use these systems, *"the decision taken about each application cannot be based exclusively on an algorithm"* and that *"it requires, on the contrary, an assessment of the application's merits by the admissions board and then by the relevant head of department."*[87]

The criterion of the need for human intervention is illustrated in this case, i.e. a review by the admissions board and the relevant head of department - people who actually have the power to alter the decision. However, since the reality and importance of this intervention is not specified, the qualification made by the Constitutional Council may, in its application, be considered by some authors to be *"open to question."*[88]

This doubt is heightened by the sheer volume of applications that the admissions board has to examine. Academic commentators[89] are sceptical about the admission board's ability to analyse applications within tight deadlines (approximately one month from the time the board accesses the applications to the time that the decision must be made).

They mention that for a law degree alone, Paris 1 Panthéon-Sorbonne University received 14,777 applications in 2018, followed by Paris 2 Panthéon-Assas University with 13,084 and Paris Descartes University with 9,841. As a senator pointed out to the Senate's Committee on Culture, Education and Communication in 2019 with regard to Paris Descartes University, a significant proportion of applications are screened out during the first ranking process by the algorithm used, and for all those applications (and the students concerned), it has to be recognised that human involvement was lacking[90]. Recently, a learned computer society explicitly went on record as stating that it was using *"an automated procedure to rank applicants."*[91]

However, the aim here is not to cast aspersions about the admissions board's members, since they are faced with a partly impossible task that sometimes means that they have no choice but to use quantitative algorithmic tools. Nevertheless, this report can only emphasise the considerable discrepancy between the law and the practices that it defines, which has an especially critical effect in this case due to the importance that the decisions for allocating places in higher education has for the scores of people concerned, many of whom are minors, every year.

THE CASE OF AFFELNET - THE ALGORITHM USED TO HELP MANAGE THE PROCESS OF ALLOCATING PLACES FOR COLLEGE STUDENTS

Every year, the online Affelnet platform enables hundreds of thousands of final-year secondary school pupils to submit their preferred college choices (whether standard or vocational colleges) and receive a place. As part of this procedure, each choice is examined against the following factors: the ranking provided by the parents, the pupil's assessments (especially their academic results) and the colleges' available number of places. Each factor is given a certain number of points, which are used to establish the scale.

An algorithm is used in each local education authority to "*facilitate the process of allocating places to pupils*"[92], according to the priorities and strategies defined by the regional director of education. In other words, this algorithm is supposed to help with the decision-making process, but not serve as the sole basis for the decision. In addition, Article D. 331-38 of the Education Code stipulates that "*Assigning places is the responsibility of regional directors of education for courses located within their region. They are assisted by a committee, whose composition and operation are defined by order of the Minister for Education*." As with the Parcoursup platform, there are indications that agents with the necessary competence and authority should be involved in the decision-making process.

However, while investigating a complaint, the Defender of Rights was able to ascertain that a decision to assign a pupil had been taken entirely automatically[93]. In this case, an assignment decision had been made for a student whose scale sheet showed only a value of "0" in the category relating to school assessments.

At the end of the first round in the AFFELNET allocation campaign in 2021, all the pupil's choices had been rejected. It was only when the pupil's parents, who were astonished by the rejections in light of their daughter's grades, asked to see the information used to take the assignment decision that the gross administrative error was uncovered and rectified. In this case, the fact that a decision was taken on the basis of a scale that showed 0.000 for the "school assessments" category, without this unusual value prompting the committee to check its accuracy, was a strong factor that led to the presumption that the decision had been taken entirely by an automatic system. The investigation by the Defender of Rights confirmed this presumption, namely that no information about the involvement of the aforementioned committee or the opinions that its members may have issued was communicated by the education authority to the institution's services.

According to the education authority, such a figure (0.000) could correspond to a case where the pupil was exempted or absent, but the incredibly unusual occurrence of such a situation should, in any event, have led the authority to check the reasons for such a figure while the decision was being considered. But this did not happen.

The fact that only one case was brought to the Defender of Rights' attention in relation to this procedure should not lead to the conclusion that this is an isolated case. On the contrary, the evidence revealed during the investigation suggests that decisions during the procedure may have been and still are entirely automated on a massive scale, at least for some pupils.

These cases may only be examples, but they raise questions about the substance of human intervention.

On the other hand, the well-documented existence of biases leads to suspicions about whether humans are truly involved in individual administrative decision-making processes.

Considering the possibility of using "significant human intervention" as the criterion for distinguishing between fully automated individual decisions and decisions taken on the basis of an algorithm is, according to the doctrine[94], part of a techno-neutral vision, which has come in for major criticism. In other words, technology, created by human will, can influence it. As philosopher Jacques Ellul pointed out in his 1977 book called *The Technological System*, "*as modern men, we are called upon not to employ technologies, but to live with and among them."* It is impossible to imagine a *"sovereign man enthroned in this collection and acting upon technology in complete independence."* In *The Technological Bluff* published in 1988, Ellul points out that to recognise that technology is not neutral is to recognise *"that it has its own weight, its own determinations, its own laws; as a system, it evolves by imposing its own logic."*

As long ago as 1956, German philosopher Günther Anders argued in his book, *The Obsolescence of Man*, that it is inaccurate to assume that technology is a means to an end. These instruments are not means, because they are decisions taken in advance, *"decisions, precisely, that are taken even before we are given the opportunity to decide"*: the *"means themselves, the instruments themselves, are not mere objects that we can use, but already determine, by their structure and their function, their use as well as the style of our activities and our lives, in short, they determine us"*.

These authors' analyses reveal the extent to which the role of human will in the use of technology is reduced. We can no longer consider that every technical object is subject to the will of a human person who would remain sovereign; on the contrary, it is vitally important to understand the way in which the operation of the technical object subjects or directs this will. This is how the so-called techno-structural position considers the complexity of the relationship between human beings and technology, and this understanding can be used to formalise, from a theoretical point of view, the criticism aimed at the criterion used to distinguish between automated and partially automated decisions.

In addition to the serious warnings about the **risks of public officials losing their critical thinking skills when faced with the results of the algorithms and AI systems that they need to deal with in a given context**[95], several types of bias must be highlighted[96]:

- **Anchoring bias**, which occurs when public officials are unable or struggle to get past the "first impression" provided by the system when that system is involved at the beginning of the process.

- **Automation bias**, which is defined as the propensity for humans to give greater credence to the results produced by the machine than their own judgement, and thereby consistently or excessively endorse the machine's recommendations.

As the Conseil d'État points out, the human approval rate for the machine's results is *"one of the main tell-tale signs of automation bias"*[97] (when the approval rate reaches 100%, there can be serious doubts about whether officials are involved). In principle, however, a statistical evaluation falls short when it comes to gauging the substance of human intervention, and only a global approach to the context and conditions of the decision-making process would appear to be relevant. In this respect, the Conseil d'État issued a proposal back in 2014[98] for *"confirming that human intervention in the decision must be real and not just formal"* and suggested *"creating a soft law instrument to specify the criteria for assessing the effective nature of human intervention."* This recommendation has yet to be implemented.

- Potential **biases caused by the liability regime**[99], which may lead the human reviewer to *"follow the algorithm's recommendations excessively"*. The reason is that making a mistake by following the algorithmic system's result or recommendation would be considered less negligent than making a mistake by rejecting the system's result or recommendation. According to Maxwell, *"this bias will be found in any situation where the human will have to justify, to her manager, for example, the rejection of an algorithmic recommendation, whereas the reverse would not be true: a decision to validate the algorithmic recommendation would not require any justification from the human decision maker."* However, the liability regime for a public official who is responsible for the decision is not clear in this context.

All these observations lead to questions about the real scope of the **distinction** between partially automated decisions on the one hand and fully automated decisions on the other. This distinction is especially important when bearing in mind that the legal framework:

- **Is based on a right not to be subject to a decision based solely** on automated processing which produces legal effects concerning him or her or similarly significantly affects him or her[100]. As seen earlier, although this is subject to **exceptions, one of which applies to the authorities' actions**, it nevertheless constitutes **a principle**.
- Has provided safeguards for fully automated individual administrative decisions, primarily the ban on using so-called sensitive personal data as the basis for fully automated individual decisions.

In addition, the way in which algorithmic processing operations contribute to decision-making is part of the information that government services must communicate to the data subject on request, when an algorithm has been used to reach an individual administrative decision, whether the decision is entirely or partially automated. As such, there is a direct link between clarifying the substance of human intervention and the information provided to users who request to see that information (see Part III).

IN LIGHT OF THESE CONSIDERATIONS, THE DEFENDER OF RIGHTS MAKES THE FOLLOWING RECOMMENDATION:

- *In cases where the individual administrative decision taken on the basis of the result of an algorithm or AI system is classed as a partially automated decision: lay down mandatory criteria and procedures, either alternatively or cumulatively, to qualify "human intervention" with a greater degree of precision.*

For example, work carried out by students at the national administration college, ENA[101], has identified a number of elements:

- The moment of human intervention
- The form of human intervention:
  > Is it aimed at endorsing or rejecting the result produced? In this case, is it a discretionary decision or a verification of the result's accuracy?[102]
  > What elements does human intervention take into account, other than the result produced by the system?

In this respect, several proposals[103] suggest defining in advance *"specific tasks for the human and for the computer",* which would be in line with the Council of Europe's recommendations on responsibility[104]. The *"precise definition of the mission of the individual human reviewer"* and the identification of the means allocated for this purpose (in terms of equipment, human resources, time, and so on) would help refocus responsibility and formalise human intervention in specific terms. Programming "effectiveness tests"[105] on human intervention would ensure its effectiveness.

These elements should be defined before the relevant government agency implements the algorithm or AI system. They would also provide a way of fulfilling the obligation laid down by the CRPA in cases where data subjects request additional information. In this case, the public service must inform the data subject of "the extent and way in which algorithmic processing contributes to the decision-making process"[106] (see Part III).

- *In the absence of criteria for qualifying human intervention, consider another method for distinguishing between partially and fully automated decisions.*

This would involve evaluating the appropriateness of basing the decision on the criterion of whether the result produced by the algorithm or AI system is "directly applicable".

The algorithmic result is said to be directly applicable when it is sufficient in itself to produce the substance of an individual administrative decision. A set of indicators is used to characterise this result, which must be:

- Clear: there should not be any need to interpret the result.

- Precise: it must be detailed and applicable to the situation under review.

- Complete: it must not leave any room for discretion.

According to Liane Huttner, who can be credited with this proposal[107], it is not the algorithm that is categorised, but its use at a given moment in time for a specific decision: consequently, the same algorithm can be qualified differently according to how it is used in the decision-making process.

On this basis, the proposal is to distinguish between:

- Decision-making algorithms that provide the complete basis for a decision.

- Decision-support algorithms that do not provide the complete basis for a decision, but may help clarify the decision.

> **The criterion for distinguishing between these algorithms consists of the "direct applicability" of the result produced by the algorithm or AI system.**

When an algorithm or AI system is used and the result is "directly applicable", the regime specified for fully automated individual administrative decisions should be automatically applied.

It should be remembered that the legal framework prohibits the authorities from taking fully automated individual administrative decisions using personal data that contain sensitive data within the meaning of the GDPR. However, looking at the example of health data alone, particularly data relating to disabilities, these data are present in a large number of processing operations, particularly in the social sector. Therefore, the scope of sensitive data needs to be examined on a European level, while taking into account the GDPR and the role of CJEU case law in this area.

In this context, the Defender of Rights recommends that the European Data Protection Board (EDPB) should establish clear guidelines on the definition of sensitive data within the meaning of the GDPR and the Data Protection Act, and that consideration should be given as to whether this prohibition is relevant in relation to the risks of discrimination. The EDPB is an independent EU body whose mission is to ensure that the GDPR is consistently applied in the EU countries.

PART III

TRANSPARENCY: A STRONG PRINCIPLE, BUT IN PRACTICE?

The transparency of government action is not a new issue and largely precedes the question of the mainstream use of AI in public services. This is evidenced by the tensions that exist between the authorities' discretionary power and the associated rights that have developed over time in favour of transparency. As Jean-Marc Sauvé pointed out[108]: *"Transparency allows the principles and values underlying public action to be effectively implemented and in so doing strengthens citizens' confidence in public institutions. It is even a cornerstone of those institutions. (…) It plays a contributory role in the effective implementation of the principle of equality. Equal access to public posts, for example, to prevent any form of nepotism or favouritism, or equal treatment for public service users to counter the privileges associated with asymmetric information. In particular, through the right of access to administrative documents and the obligation to state reasons, transparency also fosters an impartial and objective public service. This is why transparency, in its various legal forms, is a component of the right to an effective public service."[109]*

Among the decisions that are currently taken by or with support from algorithms, there are administrative decisions that were not always clearly justified and which could have been taken through opaque procedures. The use of algorithmic or AI systems can bring transparency to these very cases, because their operating rules must - at the very least - be defined.

Pursuant to Article 47 of the French Data Protection Act, the validity of an administrative decision is subject to informing users that such a decision was taken by means of an automated decision-making process[110]. However, there must be a way of identifying exactly what the nature and level of this information should be for the decision to effectively meet this obligation[111]. The question also arises as to what information should be provided about the algorithmic processing operations used in public services, which must be made more widely available to the public and end users. Notwithstanding the legal restrictions that govern such information and the practical difficulties involved, what can be done to ensure that this right to information is fully respected for public service users, and for what purpose?

1· MAJOR CHALLENGES

There have been calls to provide transparency about the automated processing operations used by government agencies and the decisions taken on this basis with the aim of bringing balance to the legal framework in a manner that is consistent with the idea that transparency is a *"form of rationalising, monitoring and above all legitimising public action."*[112] The right to information that is available to public service users whenever an algorithm is involved in making an individual administrative decision that affects them derives from the principle recognised in Article 15 of the Declaration of the Rights of Man and of the Citizen, which states that *"Society has the right to require of every public agent an account of administration."* This right also stems from the obligation to state the reasons behind certain administrative decisions, as set out in Article L. 211-5 of the CRPA[113].

Although providing relevant users with information about the algorithm used to make the individual administrative decision seems to be essential for ensuring the effectiveness of a principle of constitutional status, it should be remembered that transparency is also key to understanding certain decisions and therefore effectively discussing and challenging those decisions. As such, transparency is a vital part of ensuring accountability of the authorities' actions. In this sense, it should be seen as a prerequisite for rooting out potential errors, abuses and discrimination. Some authors refer to this issue as "fairness", a concept that encompasses transparency, but it is only one of its facets: *"Fairness presumes that the algorithm says what it is doing, but also that it is doing what it says and that what it says corresponds to satisfying the general interest, in compliance with the law (...). Basically, fairness means that the algorithm must be transparent, honest and legal."*[114]

Therefore, the mainstream use of algorithms in public services is helping renew the requirement to ensure transparency in public action.

A distinction can be made between several phases, with each phase helping make the right to information a reality[115]:

- Information is aimed at "making visible", showing and bringing to the user's knowledge.

- Explanations are aimed at "making the decision understandable" - note that a distinction can be made between specific explanations (to explain a particular decision or result) and global explanations (relating to the general logic behind the system).

- Justification is aimed at "making the decision acceptable" - therefore, it draws on external elements, such as statutes and public action objectives, to justify the result[116].

As will be seen, **some transparency information is intended solely for the benefit of the user concerned, but other obligations target a wider audience**. A prime example is the obligation for certain government agencies to upload the rules defining the main algorithmic processes that they use to perform their duties when these processes are used for individual decisions. The challenge of providing readily understandable information to single users is matched by the challenge of ensuring that public information is accessible, so that various stakeholders[117] can adopt it, participate in understanding the impact that these systems have on society[118] and *"understand, challenge and discuss the role that machines play in public decision-making."*[119] Such collective mobilisation would help users fully understand the issues, bearing in mind that they almost never possess all the skills required to assess the algorithm. This is despite the fact that the ability to assess the algorithm is essential for identifying any questionable or even discriminatory effects, which would only be apparent on a collective scale, based on an analysis of the source code and/or aggregated results produced by the system[120].

However, what constitutes a right for users and the stakeholders involved in these issues also constitutes, for the authorities themselves, **a challenge for understanding their own actions**[121] or even for taking (back) control of the tools and therefore their sense of accountability.

It would appear to be essential to "*give the authorities a central role in assessing the algorithms used*" and prevent them from "*using tools without understanding how they work, which would otherwise make them dependent on the results produced by their algorithms.*"[122]

This is a fundamental issue at a time when public services are ramping up their use of such IT systems and often relying on private service providers, and when *"the rate at which algorithms need to keep pace with changes in legislation and regulations is a destabilising factor for employees working in information systems departments."*[123] Although the jurisdiction of the Defender of Rights concerns the rights of public service users and not the general operation of these services[124], the two appear to be linked. The idea is to ensure that users' rights are respected through its internal operation and by developing officials' ability to understand the systems used. As Liane Huttner points out, the person behind the decision is just as concerned as the decision's recipient by the various mechanisms designed to subject the algorithm and the decision taken to a process that is both controlled and open to challenge. Officials must be capable of understanding how the system works and stepping in to review the results[125].

For example, the transparency of high-quality source code, i.e. without any data masking[126], should also be a "desirable objective (...) for the authorities and the IT department concerned in the first instance, since these techniques contribute to good software engineering practices and the maintenance quality of the information systems."[127] This objective must be compared with the obligation (under certain conditions) to publish the source code used by the authorities.

INSIGHT

THE ISSUE OF EXPLAINABLE AI OR "XAI" SYSTEMS

In general, AI systems and algorithms are sometimes clouded in ambiguity, and their results can be hard to interpret. However, as stated by the EU AI Act[128]: "*the exercise of important fundamental procedural rights, such as the right to an effective remedy and to a fair trial, as well as the rights of defence and the presumption of innocence, could be hampered, in particular, where such AI systems are not sufficiently transparent, explainable and documented*."

As defined by the CNIL[130]: "*In the field of artificial intelligence, explainability is the ability to help users consider and understand the elements that the AI system uses to produce a result. For example, this could be the input variables and their impact on predicting a score, and therefore on the decision. Explanations must be tailored to the intended recipient's level of understanding.*"

There is nothing new about explainability in the field of AI. Nevertheless, as some authors[130] have pointed out, it has become a central issue in the wake of the development and importance of so-called "black box" systems. Note that, according to the same authors, the use of this expression refers to two distinct cases. Firstly, there is the instance of a so-called "proprietary" system where there is no access to the source code for people without a licence. Secondly, there is the instance of a system that is so complex that it cannot easily be understood.

In recent years, a global research community has emerged to address the theme of explainable AI (known as xAI)[131].

Daniel Le Métayer and Clément Hénin[132] are just two of the many researchers who are taking a closer look at the various ways of tackling opaque algorithmic systems.

Many different methods are being examined, but no consensus has yet been reached, as noted by the CNIL[133]. Some work in a "black box", i.e. without any knowledge of the system code, while others work in a "white box", i.e. they take action on the code. Some aim to make the systems interpretable (or "intelligible"), while others endeavour to produce post hoc explanations for opaque systems.

Although each approach has its advantages and disadvantages, a number of challenges stand in the way.

Provide explanations or justifications that can genuinely be understood by lay users (doctors, lawyers or people affected by the system's results, for example)[134].

In general, the explanations provided by existing methods are one-way, meaning that they are given by the responsible party to the person concerned without any discussion and delivered in a fixed, predefined form (e.g. decision trees or lists of overriding factors), whereas users have varied needs depending on their motivations and level of proficiency.

To date, only scant research has been conducted into the justifications and challenges. An experimental evaluation of the results is an issue that warrants greater attention. It is easier to define KPIs for the system and assess the system's performance than explain it and measure the quality of the explanation[135].

2· RIGHTS ALREADY IN FORCE

Two separate legal frameworks set out requirements governing the transparency of fully or partially automated individual administrative decisions.

Firstly, there is the legal framework for protecting personal data. Compliance with its requirements is monitored primarily by the CNIL[136].

Secondly, there is the legal framework governing relations between the public and the authorities. Cada is the authority responsible for ensuring compliance with the freedom of access to administrative documents and public records, and the re-use of public information[137], but it has no jurisdiction to oversee compliance with certain transparency-related rights mentioned in the CRPA Code.

Public service transparency extends beyond the subject of the algorithms and AI systems used for citizens: the right of access enshrined in the CRPA concerns all types of documents (cf. Article L. 300-2 of the CRPA).

On this subject, proposals to reform the system are emerging in public debates about the right of access to administrative documents in general, such as extending the deadline (currently one month) that the authorities have to respond to access requests before the case can be referred to Cada for an opinion, transforming Cada's opinions into approved opinions that would be binding on the authorities (which is currently not the case), and increasing its budget. This report is focused on the use of algorithmic systems in public services, so it is not aimed at comprehensively addressing these issues.

As far as the applicable legal framework is concerned, it is important to understand that:

- The "personal data" framework applies exclusively to decisions (a high number in real life) that are taken (by the government agency or private company) on the basis of personal data processing and, in the context of algorithms in the public sector, and that its provisions relating to explanations about the logic behind the algorithm apply only when the decision is fully automated or profiling is established within the meaning of Article 22 of the GDPR[138].

- The "CRPA" framework applies more broadly: as soon as an administrative decision is involved (whether fully or partially automated, whether or not based on personal data processing) and therefore, in many cases, cumulatively with the "personal data" framework.

- The content of the respective obligations is not the same, but broadly similar.

The following summary tables provide an overview of the obligations arising from both legal frameworks, insofar as they relate to the transparency of partially or fully automated individual administrative decisions.

ACCORDING TO THE CRPA

> Whether the individual administrative decision is partially or fully automated, whether <u>or not</u> it is based on processing personal data, and whether or not it concerns a natural person <u>or a</u> legal person.

· **Information to be provided to the person concerned by the individual administrative decision:**

This is a two-step mechanism:

First step[139]: the individual decision must contain an **explicit statement informing the person concerned that the said decision has been taken on the basis of an algorithm** and that they may request additional information.

! Exception if it violates protected secrets[140].

Second step[141]: if the person concerned by the decision requests further information, the authorities must provide four sets of information in an intelligible form relating to the algorithmic processing operation that was used as the basis for the decision (and to the extent permitted by protected secrets):

1. The extent and way in which algorithmic processing contributes to the decision-making process.
2. The data processed and their sources.
3. The processing parameters and, where applicable, their weighting, applied to the situation of the affected person.
4. The operations carried out by the processing.

· **General information to be published online on the website of the relevant government agency:**

In pursuance of Article L. 312-1-3 of the CRPA, government agencies[142] with over 50 FTEs[143] must upload the rules defining the main algorithmic processes used in the performance of their duties when they form the basis of individual decisions.

An exception applies if this violates protected secrets[144].

· **Access to and publication of the source code via a "Cada" request:**

The CRPA specifies that the source code[145] constitutes an administrative document[146].

As such and in the same way as minutes, reports, studies, and so on, it forms part of the information that <u>any person</u> may, within the framework laid down by this CRPA Code, request from the State, regional or local authorities, or other public or private legal entities entrusted with a public service mission, as part of a so-called "Cada" request[147].

The source code must also be published online if such a request is made to the authorities[148] and ultimately by default[149] (the last obligation only applies to local authorities with over 3,500 inhabitants and government agencies with over 50 FTEs[150]).

· **Generally, access to an administrative document concerning the algorithm used by means of a "Cada" request:**

The traditional right of communication can be used to obtain any administrative document that prepares or assesses the results of the algorithm (unless it is a protected secret).

ACCORDING TO THE GDPR AND FRENCH DATA PROTECTION ACT

> When the administrative decision is fully automated <u>and</u> based on personal data processing <u>and</u> concerns a natural person

Transparency arrangements:

Article 12 of the GDPR sets out the details of the information that must be provided to data subjects (information to be provided in a concise, transparent, intelligible and easily accessible form).

Information to be provided to the data subject:

First step: depending on the case, Article 13 or Article 14 of the GDPR sets out the content of the information to be provided to data subjects (and not at their request) about the processing operations on their personal data.

In the case of a fully automated decision, including profiling, the controller must provide the data subject with meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject (Articles 13(2-f) and 14(2-g) of the GDPR).

Second step: at the data subject's request, the controller must *"be able to explain to the data subject, in detail and in an intelligible form, the manner in which the processing operation has been carried out in relation to him or her"* (Article 47 of the Data Protection Act).

Third step: data subjects may also exercise their right of access[151] (confirmation that their personal data are being processed and a certain amount of information about this subject).

Requests for access to the records of processing activities:

Barring exceptions, each controller must maintain a record of processing activities containing a certain amount of mandatory information[152] (purposes of the processing, description of the categories of personal data, categories of personal data recipients, etc.).

This record of processing activities, which can be used to understand the algorithms involved, can be requested by means of a so-called "Cada" request.

Therefore, it is clear (especially from the provisions of the CRPA) that the obligations laid down are considered from both **an individual point of view** (for the user concerned) and **a collective point of view** (for such third parties as associations, research teams, people accompanying users, and any interested citizens or groups of citizens).

INSIGHT

TRANSPARENCY REQUIREMENTS AND THEIR CONSEQUENCES FOR THE ALGORITHMS THAT MAY BE USED FOR INDIVIDUAL ADMINISTRATIVE DECISIONS

· In a 2018 ruling on **fully automated individual administrative decision-making**, the French Constitutional Council considered that *"the controller must have control over the algorithmic process and its changes, so that it can explain to the data subject, in detail and in an intelligible form, the manner in which the processing operation has been carried out in relation to him or her. As a result, algorithms may not be used, as the sole basis for an individual administrative decision, where they are capable of revising the applicable rules themselves, without oversight and validation by the data controller."[153]* Therefore, the Constitutional Council has taken the transparency requirement as its starting point for **imposing a strong limitation on systems that are likely to serve as the basis for fully automated decisions**.

· It would appear that this limitation can be **extended to systems used to make partially automated decisions**. The information that the public sector must provide to data subjects at their request when a system has been

used (partially or fully) to establish a decision includes, as we have seen, *"the processing parameters and, where appropriate, their weighting, applied to the situation of the affected person."[154]*

- According to the CNIL, *"the parameter is the learned property of the data used for training (for example, the weight of each neuron in a network)."[155]*

- However, it is not certain that the exact weighting of the parameters in a model derived from machine learning can be accurately determined. According to legal expert Winston Maxwell, this *"will remain unclear."[156]* While he considers that certain explainability methods can help *"guess the approximate weights of the different features,"* he emphasises that these tools provide only an approximation.

> The direct consequence of this observation is that using learning algorithms as a basis, even in part, for individual administrative decisions must be considered in light of the applicable regulatory provisions, including the requirement for transparency.

In terms of how regulations have changed, the EU AI Act explicitly provides for a "right to explanation of individual decision-making" for any person concerned by a decision taken by an organisation deploying an AI system considered to be high-risk and which produces (legal or significant) effects on that person[157]. In this case, the affected person has the right to obtain *"clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken."*

In addition to the fact that the AI Act makes **no distinction here between fully automated** and partially automated decision-making, it is worth noting that these provisions in the EU AI Act refer to a criterion of "meaningful", without actually specifying at this stage what that means.

3· LEGAL BOUNDARIES: A RECOMMENDATION FOR IMPROVEMENTS

From a legal perspective, the boundaries on transparency and the communication of information relate to:

1. The fact that **the applicable legal framework is sometimes twofold**: in the case of a fully automated individual administrative decision taken on the basis of personal data processing, the organisation making the decision must comply with the information requirements in the CRPA and also the requirements in the Data Protection Act.

THEREFORE, THE DEFENDER OF RIGHTS MAKES THE FOLLOWING RECOMMENDATION:

- The public authorities should consider simplifying the transparency obligations for organisations in an effort to improve the information provided to the users concerned without reducing their rights.

2. *Protect certain secrets:*

- This may include information that cannot be disclosed in the areas directly listed in the CRPA[158], especially *"public security and personal safety"* or *"the security of government information systems"* or *"the investigation and prevention, by the competent services, of offences of any kind"*[159]; Cada is responsible for assessing what constitutes a secret and how it applies to the case in question:

  > For instance, when it comes to secrets relating to the "security of government information systems", Cada states that it assesses *"this reservation very strictly, since it cannot be presumed and must be demonstrated from detailed information provided by the authorities."*[160]

  > Cada rejected a request to disclose the algorithm of the data mining model that CNAF, the national family allowance fund, uses to assign a risk score to beneficiaries, on the grounds that it would undermine the organisation's policy of tackling social fraud and impact the effectiveness of its inspections. It considered that disclosing the variables involved and the associated coefficients would be tantamount to divulging the criteria used to target inspections, which would heighten the risk of individual or organised fraud. However, Cada approved a request to disclose the variables in models that were no longer in use and which were no longer involved in current or future inspections[161].

- This may include information listed in special laws, such as the case with the Parcoursup system: to *"protect the confidentiality of the deliberations by the educational teams responsible for examining applications,"* Article L. 612-3 of the Education Code excludes the application of two articles in the CRPA that relate to the disclosure and publicity of the algorithmic processes used as the sole or partial basis for individual administrative decisions, as noted and endorsed by the Constitutional Council in its decision on the matter[162].

3. *Respect other rights, in particular:*

- **The right to privacy**: a number of documents may only be disclosed to the interested party[163].

  > Cada rejected a request to disclose the trained learning models of the tool developed by the Court of Cassation for pseudonymising court decisions. It considered that granting this request would undermine the very purpose of the tool, i.e. tighten up security measures concerning the public dissemination of court decisions while respecting the privacy of the individuals concerned, due to the risk of operations that could piece together the hidden data. In this respect, Cada considered that disclosing these documents could potentially infringe the protection of privacy[164].

- **Intellectual property rights**: the source code of an algorithm may be an administrative document that can be disclosed, but in practice, intellectual property regulations may prevent its disclosure where the algorithm has been developed as part of a system covered by a public contract. According to the Conseil d'État, a document that does not belong to the public authorities may only be disclosed with consent from the person holding the ownership rights[165].

4. *Non-binding nature of certain procedures:*

- The law does not provide for any **penalties if the authorities and other organisations concerned**[166] **fail to comply with the aforementioned transparency obligations** in the CRPA, except where the decision is held to be invalid in the absence of the required statement about the use of an algorithm in the decision-making process. In particular and barring the exceptions relating to protected secrets, the obligation to publish the "rules defining the main algorithmic processes" within the meaning of Article L. 312-1-3 of the CRPA is not subject to any penalties.

- **The non-binding nature of the opinions issued by Cada**: when a request for access to an administrative document is referred to Cada after it has been rejected or overlooked by the authorities, Cada may issue an opinion in favour of disclosure, but cannot impose it on the authority concerned. To obtain access to an administrative document, the matter must sometimes be referred to the relevant administrative court.

4· PRACTICAL LIMITATIONS AND RECOMMENDED REMEDIAL MEASURES

From a practical perspective, implementing the right to information set out in the CRPA and the principle of transparency runs into various limitations and difficulties. For each case, the Defender of Rights issues one or more recommendations to ensure that the right to information for public service users is put into practice.

SEVERAL TYPES OF LIMITATIONS HAVE BEEN IDENTIFIED:

1. Failure to publish the "*rules defining the main algorithmic processes*".

Ever since the **obligation under Art. L. 312-1-3 of the CRPA** came into force, several stakeholders[167] still claim that there are "doubts surrounding its effectiveness", and some even go so far as to say that it has generally remained a "dead letter" since 2016[168], with a few exceptions[169].

The lack of penalties for failure to fulfil this obligation may contribute to the lack of enthusiasm shown by the government agencies concerned. If the algorithmic rules were published, such as by different local authorities of the same type (e.g. regions), it could help stakeholders compare and discuss their practices and the policy choices made, which are not always explained elsewhere.

The Defender of Rights welcomes the move to establish local and national citizens' rights watchdogs (including the "Ouvre-Boite" association[170] and the ODAP[171] public algorithm observatory project), which should not, however, conceal the responsibility of the relevant government agencies in this respect.

THEREFORE, THE DEFENDER OF RIGHTS MAKES THE FOLLOWING RECOMMENDATION:

· *The Government should introduce a penalty for failure to fulfil this obligation and set up a nationwide survey relating to compliance with these publication obligations.*

· *The government agencies and organisations concerned should promptly comply with the publication requirements stipulated in the CRPA.*

- *At the same time, set up and support local citizens' rights watchdogs or surveys on a national scale.*

2. **Difficulties** within public services in governing the application of the legal framework, involving a wide range of stakeholders, often without any one necessarily being designated as responsible (data protection officer (DPO) - person responsible for access to administrative documents (PRADA), a role that is sometimes little known[172] - ministerial administrators of data, algorithms and source code (AMDAC) and **practical difficulties for the public authorities, local authorities and other organisations concerned in complying with existing obligations**, which implies:

- **Defining an individual statement** about the use of algorithms, which should be affixed to the individual administrative decision concerned within the meaning of the CRPA, which is clear and accessible to users.

- **Knowing which further information** must be provided to the person concerned in response to an additional request under the CRPA.

- **Including this information in a clear, intelligible and understandable form** (design issues).

IN LIGHT OF THE GOVERNMENT AGENCIES' NEED FOR SUPPORT AND CLARIFICATION, THE DEFENDER OF RIGHTS RECOMMENDS THAT THE GOVERNMENT SHOULD:

- *Deliver effective support to the public sector, local authorities and other organisations concerned in response to the lack of expertise and incentives (due to the absence of penalties for certain obligations), and the challenge of facilitating dialogue between the structures concerned, and users, and more generally fostering public debate.*

- *Determine which algorithmic and AI systems may be used by the public authorities when serving as the basis for a partially automated individual administrative decision, so that compliance is ensured with the CRPA's transparency obligations* (specifically those relating to the disclosure of the processing parameters and, where appropriate, their weighting, applied to the situation of the affected person)*.*

- *Concerning the guide to public algorithms proposed by Etalab[173], which contains details and an example of an explicit statement on the use of algorithms:*

  - Enrich it with examples of additional information, better promote the guide,

and increase the resources of Etalab or a dedicated structure[174] for this purpose.

- Add information about the *"rules defining the main algorithmic processes used"*, as well as examples.

- Adapt it, where applicable, for local and regional authorities (particularly in terms of examples) and obtain the endorsement of the Ministry for Local and Regional Authorities and the National Agency for Territorial Cohesion (ANCT).

3. One of the difficulties facing government agencies on the one hand and citizens on the other is caused by the fact that the request for additional information from the person concerned (by the individual administrative decision) sometimes fails to explicitly mention the relevant articles of the CRPA[175]. This can make it hard for the authorities to identify the terms of the user's request [176].

THEREFORE, THE DEFENDER OF RIGHTS MAKES THE FOLLOWING RECOMMENDATION:

• *Government agencies and other relevant organisations should set up in-house training for their employees to ensure that requests for additional information are identified by those receiving them, especially through examples.*

• *Based on the tools already developed by associations[177], the Government should:*

- *Produce templates for users to submit their requests* (for additional information and for requests to publish the source code).

- *Make such templates visible and easily accessible on the websites of the government agencies and other organisations concerned* (including Cada).

4. Finally, there are undeniable difficulties caused by the technical nature of the systems and the varying levels of proficiency possessed by the people involved. According to a survey in 2023, only 24% of respondents claim to know exactly what AI means [178]. In the same vein as the "digital divide"[179], this is termed "opacity as technical illiteracy"[180], combined with a relative lack of awareness of the right to information and the right of access. In this sense, even if and when the government agencies concerned comply with the legal obligations[181], this is not enough to meet the transparency requirement, because the information and documents published are still essentially inaccessible to the public[182], who bear the burden of trying to understand them. Faced with a fully or partially automated individual administrative decision, information can be considered meaningful if it improves understanding for users and allows the users concerned or the group defending them to challenge the decision taken[183].

THEREFORE, THE DEFENDER OF RIGHTS MAKES THE FOLLOWING RECOMMENDATION:

• *To the Government:*

- Taking into account the Constitutional Council's considerations in 2018[184] and the provisions of the EU AI Act relating to the right to explanation of individual decision-making (but only for decisions taken through a high-risk AI system), *enshrine a "right to explanation" for all partially and fully automated individual administrative decisions[185].*

- Taking into account the technical nature of the subject, *support research efforts in this area as well as collective and association-led projects aimed at improving understanding and promoting public debate on these issues*, especially since the effects of these systems can also be felt on a collective scale - in connection with the risks of discrimination which, in algorithmic logic, can very often only be identified on this scale.

· *To the government agencies concerned:*

 - In general, and in keeping with the recommendations issued by the Conseil d'État[186], *involve public service users at every level*, i.e. from the moment that the choice is made to use algorithmic processing or an AI system, in preparing the wording of the information, in the forums set up, where applicable, during the system's evaluation, and within the ethics committees where such committees exist.

 - *To ensure that this right to explanation becomes as much a reality as possible, take advantage of the existing tools and develop new systems*.

   To make sure that the administrative decision-making system can also be used to provide personalised explanations about the algorithmic basis behind the decision:

   · The technical system for explaining the decision should be aligned with the technical system for making that decision, and the decision-making system must be supported by the legal elements that underpin and justify the decision[187].

   · Assess the feasibility of automatically producing explanations that are personalised, detailed and intelligible to those involved in the internal and external review of the decision-making system[188].

   · Use the Algocate[189] solution to automate the "justifications" component.

   To make it easier for the users concerned to understand the explanations provided[190], bearing in mind that access to the algorithm alone is insufficient for understanding the result that led to a decision[191]:

   · Develop explanatory tools that present counter-arguments and highlight the limitations of algorithmic predictions.

   · Propose gaming strategies to encourage people to understand the information presented and search for relevant information more effectively.

   · Provide tools that allow relevant users to test the algorithm by changing the input data, such as their age or family quotient[192].

It would appear that the efforts made to **scale up the use of algorithmic systems across government services should be accompanied by efforts to fulfil the obligations laid down by law and the regulation[193], work on explainability and implement existing doctrinal proposals**, so that the rights that are already in force for public service users can be put into practice, while implementing a genuine right to explanation of fully or partially automated individual administrative decisions.

It is only once this condition is met that transparency will be able to *"recover its democratic dimension,"* as Jean-François Kerléo[194] points out, noting that the ongoing process of technicalising the transparency issue is characterised by *"an accumulation of practices that are increasingly reserved for the initiated."*

CONCLUSION

Five years after fully automated individual administrative decisions were introduced, the mainstream use of algorithms across the public sector raises a number of questions about respect for users' rights. The Defender of Rights is particularly interested in so-called partially automated decisions and the substance of stated and actual human intervention, and also in the legal and practical issues surrounding the transparency of the systems and decisions taken.

At a time when increasingly sophisticated automated systems are currently or on the verge of being rolled out across the public sector, the answers to these questions are especially important, because they form the cornerstone for building users' confidence in public services. The recommendations in this report are intended to play a contributory role in reinforcing users' rights and ensuring the quality and legitimacy of public services.

To achieve this aim, it is vitally important to give (back) the authorities a central role in assessing the AI systems used, by ensuring that they have the requisite in-house skills and fostering dialogue between teams in different agencies, so that the authorities are in full control of the choices that underpin the operation of such systems and the resulting individual decisions, while making sure that all the issues raised by these systems are properly understood within the public service concerned.

It would also appear to be essential to aid research in this area, and support the emergence of groups and the associations engaging with these issues. Civil society has already started taking action in a number of areas, including AI systems deployed in public spaces using algorithm-driven CCTV systems[195] and AI systems used to tackle social security fraud.

Finally, analysing the aggregated results is key to challenging the objectives pursued and redirecting the system: only an analysis of the collective and structural effects of these systems can usefully inform public debate. In light of the issues raised by these objects, understanding the collective dimension of their effects is the only way to reveal their potential infringements on users' individual rights and, ultimately, help enforce such rights.

NOTES

[1] See Defender of Rights, Dematerialisation of public services, 2019; Defender of Rights, Dematerialisation of public services: three years later, where are we? 2022.

[2] Defender of Rights, CNIL, Algorithms: preventing automated discrimination, 2020.

[3] Defender of Rights, Biometrics: the urge to safeguard fundamental rights, 2021.

[4] See CNIL, How can we enable people to stay in control? The ethical challenges of algorithms and artificial intelligence, 2017.

[5] See Conseil d'État, Artificial intelligence and public action: building trust, serving performance, 2022; Conseil d'État, Annual study 2014, Digital technology and fundamental rights, 2014.

[6] From a wealth of literature, including: Denis Merigoux, Marie Alauzen, Louis Gesbert et al., From algorithmic transparency to automatized explainability: understanding IT, legal and organizational challenges, Research report no. 9535, INRIA, Paris, 2024, p.68; ENA, Promotion 2018-2019 MOLIERE, Ethics and responsibility for public algorithms, Collective report commissioned by a central government agency at the request of Etalab, 2019.

[7] From an especially rich corpus of literature, including: Observatoire Data Publica, Justine Banuls (principal author), Transparency of public algorithms, Les cahiers de l'observatoire Data Publica, 2023.

[8] Defender of Rights, Annual Activity Report 2023, 2024.

[9] Ibid.

[10] Defender of Rights, Dematerialisation of public services: three years later, where are we? 2022.

[11] French Data Protection Act no. 78-17 of 6 January 1978.

[12] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation - GDPR).

[13] Law no. 2016-1321 of 7 October 2016 for a Digital Republic, known as the "Lemaire Law".

[14] For example, in the field of education, regarding the Parcoursup admissions procedure: see Article L. 612-3 of the Education Code, as amended by Law no. 2018-166 of 8 March 2018 on student guidance and success.

[15] Regulation (EU) 2024/1689 of 13 June 2024, laying down harmonised rules on artificial intelligence.

[16] Annex III of the EU AI Act lists, by type or area, AI systems considered to be high-risk. For a complete overview of the systems in this category, also refer to Annex I of the text, in accordance with Article 6 of the Act.

[17] Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law, adopted by the Ministers of Foreign Affairs of the 46 Member Countries of the Council of Europe on 17 May 2024 in Strasbourg.

[18] CNIL, Cada, Etalab, Practical guide to the online publication and reuse of public data (open data), 2019, p. 26.

[19] Ibid.

[20] Michel Blancard, Opening the code for calculating income tax, years 2010 to 2015, at https://www.etalab.gouv.fr/, published on 17 September 2017.

[21] Machine learning algorithms learn through training: an algorithm first receives examples whose results are known, then makes a note of the difference between its predictions and the correct results, and finally uses weightings to refine the accuracy of its predictions until they are optimised. By giving computers the "ability to learn", the complex process of optimisation (monitoring and analysing the variables in the available data to generate accurate predictions about the future) is carried out by the algorithm, which can even be left to specify the characteristics that should be considered first.

[22] According to Futura Science, "Deep learning is based on a network of artificial neurons inspired by the human brain. This network comprises dozens or even hundreds of "layers" of neurons, each of which receives and interprets information from the previous layer. For example, the system will learn to recognise letters before tackling the words in a text, or determine whether there is a face in a photo before discovering who it is."

[23] Regulation (EU) 2024/1689 of 13 June 2024, laying down harmonised rules on artificial intelligence, Art. 3.

[24] CNIL, How can we enable people to stay in control? The ethical challenges of algorithms and artificial intelligence, 2017, p. 51.

[25] Boris Barraud, "Algorithmization of government services", Revue Lamy - Droit de l'immatériel, 2018, no. 150, pp. 42-54.

[26] Elisabeth Grosdhomme Lulin, Governing in the big data era: promises and perils of algorithmic public action, Institut de l'Entreprise, 2015, 78p.; Gilles Drouet, Algorithms and public decisions, CNRS, 2019.

[27] See Dinum, DINUM Roadmap - A digital strategy for effective public action at https://www.numerique.gouv.fr/, published on 9 March 2023.

[28] See Etalab, Guide to public algorithms at https://etalab.github.io/algorithmes-publics/guide.html.

[29] In this respect, see: Cour de Cassation, Anonymised decisions: two data scientists at the Cour de Cassation at https://www.courdecassation.fr/, published on 15 January 2020: "The Court of Cassation, in liaison with the Ministry of Justice, wants to develop machine learning techniques with the aim of identifying the data

to be anonymised in court decisions before they are made accessible and available for reuse."

[30] See Law no. 2023-380 of 19 May 2023 on the 2024 Olympic and Paralympic Games and on various other measures, which authorises the use of algorithms on a trial basis to detect several types of events in order to ensure security and safety at sports, recreational or cultural events which, due to the scale of their attendance or their circumstances, are especially exposed to the risk of terrorist acts or serious threats to personal safety. These events are provided for in French Decree no. 2023-828 of 28 August 2023 and are as follows: presence of abandoned objects, presence or use of certain weapons, failure by a person or vehicle to respect the common direction of traffic, crossing or presence of a person or vehicle in a prohibited or sensitive area, presence of a person on the ground following a fall, crowd movement, excessive density of people, and outbreaks of fire. These events can be detected by applying AI systems to CCTV and drone footage recorded in public spaces.

[31] Improved medical diagnosis and prevention using medical imaging and risk factor modelling algorithms. French Decree no. 2024-468 of 24 May 2024 on the implementation of the digital health space (ENS) defines the conditions for using the data contained in the ENS for the purposes of personalised prevention for data owners and opens up the possibility for users to submit health questionnaires and receive prevention information in their ENS.

[32] In this respect, see: Emile Marzolf, The national education authorities step up the use of AI in every class, ranging from 6 to 15-year-olds, at https:// acteurspublics.fr/, published on 23 January 2024. The article goes on to say that this is especially the case with "Adaptativ'Maths", which is designed to provide teachers with "assistance based on an adaptive pathway for groups of pupils, involving the use of a dynamic personalisation algorithm", or Lalilo, which is intended to help pupils "improve their oral comprehension and pronunciation through its speech synthesis and recognition capabilities."

[33] Mirroring the efforts that the authorities in Nantes have been leading since 2020, the Île-de-France region is considering implementing a system in 2024 for anticipating the number of meals to be delivered every day to the public colleges under its authority (around 450).

[34] Prior to the adoption of Law no. 2016-1088 of 8 August 2016 on employment, modernising social dialogue and securing career paths, its effects were assessed using software (Worksim) that models and simulates the French labour market.

[35] CNIL, Artificial intelligence: the CNIL action plan, at https://www.cnil.fr/fr, published on 16 May 2023.

[36] See Info Gouv, AI: have you heard of Albert? at https://www.info.gouv.fr/, published on 22 May 2024.

[37] ENA, Promotion 2018-2019 MOLIERE, Collective report commissioned by a central government agency [Etalab], Ethics and responsibility of public algorithms, 2019, p. 1.

[38] See Marianne Uguen, The use of algorithms in public action, Prof. Marcel Morabito (ed.), master's thesis, political science, Sciences Po - School of Public Affairs, 2023, p. 12.

[39] Jacques Thyraud, Report on behalf of the Committee on Constitutional Law, Legislation, Universal Suffrage, Regulation and General Administration, on the bill adopted by the National Assembly relating to data processing and freedoms, Senate, 1977, p.22.

[40] French Data Protection Act no. 78-17 of 6 January 1978, Article 47: "No court decision involving an assessment of a person's behaviour can be based on the automated processing of personal data intended to assess certain aspects of that individual's personality."

[41] Bill no. 490, 13 December 2017, on personal data protection, p. 14.

[42] Law no. 2018-493 of 20 June 2018 on personal data protection.

[43] Personal data protection bill, Impact assessment, 12 December 2017.

[44] French Data Protection Act no. 78-17 of 6 January 1978, Article 47: "No decision producing legal effects on a person or significantly affecting him or her may be taken where based solely on automated personal data processing, including profiling (...)."

[45] Regulation (EU) 2016/679 of 27 April 2016, GDPR, Art. 22.

[46] Law no. 2018-493 of 20 June 2018 on personal data protection, Art. 21.

[47] In this respect, refer to the CNIL, "Law Enforcement Directive: what are we talking about?", 20 February 2019.

[48] Constitutional Council, 12 June 2018, no. 2018-765 DC.

[49] Defender of Rights, CNIL, Algorithms: preventing automated discrimination, 2020.

[50] Nicolas Kayser-Bril, Austria's employment agency rolls out discriminatory algorithm, sees no problem, at https://algorithmwatch.org/en/, published on 6 October 2019.

[51] Impact assessment, Personal data protection bill, 12 December 2017, p. 64.

[52] CJEU, 1 August 2022, no. C-184/20; CJEU, 5 June 2023, no. C-204/21; CJEU, 4 July 2023, no. C-252/21; CJEU, 4 October 2024, no. C-21/23.

[53] See for example: CNIL, deliberation no. 2023-134, 7 December 2023 on the bill amending French Decree no. 2003-485 of 5 June 2003 on the population census: "The case law of the Conseil d'État tends to consider that the sole information according to which a person is affected by a disability, without specifying the type and extent, would not constitute health data within the meaning of the GDPR (CE, 10th and 9th sections combined, 19 July 2010, no. 317182, Reports; CE, 10th - 9th sections combined, 28 March 2014, no. 361042; CE, 10th - 9th chambers combined, 10 June 2021, no. 431875, Reports). Therefore, insofar as the "Gali" indicator does not necessarily reveal a disability and does not directly provide any information about the type or severity of a potential disability, it could be considered that it is not health data under the terms of the aforementioned case law. In any event, the CNIL will

take special care in ensuring that the dissemination of this information excludes any data that can be used to re-identify individuals or stigmatise vulnerable groups of people, and also special attention to future data interconnections and reconciliations that could lead to the data collected under the "Gali" indicator being reclassified as sensitive data within the meaning of the GDPR."

54 According to Law no. 2008-496 of 27 May 2008 containing various provisions for aligning with Community law in the field of anti-discrimination, the list of protected grounds is as follows: origin, sex, marital status, pregnancy, physical appearance, particular vulnerability resulting from their economic situation, apparent or known by its author, surname, place of residence or bank account, state of health, loss of autonomy, disability, genetic characteristics, morals, sexual orientation, gender identity, age, political opinions, trade union activities, ability to express themselves in a language other than French, actual or assumed membership or non-membership of a particular ethnic group, nation, alleged race or religion.

55 CJEU, 7 Dec. 2023, no. C-634/21.

56 Cf. La Quadrature du Net, Scoring beneficiaries: the indecent and now undeniable practices of the national family allowance fund, at https://www.laquadrature.net/, published on 27 November 2023; Gabriel Geiger, Soizic Pénicaud, Manon Romain, Adrien Sénécat, Profiling and discrimination: investigation into the abuses of the national family allowance fund's algorithm, at https://www.lemonde.fr/, published on 4 December 2023.

57 Constitutional Council, 12 June 2018, no. 2018-765 DC.

58 CNIL, How can we enable people to stay in control? The ethical challenges of algorithms and artificial intelligence, 2017.

59 French Regulation of 17 October 2018, amending the French Regulation of 6 November 1996, as amended, approving the rules for distributing and allocating organs from a deceased person with a view to transplantation, Art. 1, 2.2, NOR: SSAP1828392A.

60 CRPA, Art. L. 312-1-3 - for government agencies with over 50 FTEs and subject to protected secrets (cf. CRPA, Article L. 311-5).

61 Soazig Le Nevé, The in-house algorithm at Sciences Po Bordeaux favours college students with grants and students from the "Cordées de la Réussite" scheme, at https://www.lemonde.fr/, published on 23 May 2023.

62 In this case, the amended French Regulation of 31 December 2020 creating an automated personal data processing system called "Parcoursup", NOR: ESRS2035793A.

63 See in particular Vincent Tiberj, Parcoursup or the algorithm-based selection process, at https://laviedesidees.fr/, published on 12 January 2021.

64 Lucie Cluzel-Métayer, "Transparency and fairness in public algorithms", in Jérémy Bousquet, Thibault Carrère, Sabrina Hammoudi, Algorithmic public action: risk and outlook, Mare & Martin, Paris, 2023, pp. 167-180.

65 A false positive is the wrong detection of a condition tested during an evaluation process (for example, if the AI system incorrectly identifies a text as written by an AI system, when the text was actually written by a human).

66 A false negative is an AI system's failure to detect a condition tested during an evaluation process (for example, an AI system fails to detect the fact that a person has fallen to the ground, even though this has actually happened and the system is supposed to identify every time that such an incident occurs).

67 Article 3(2) of the EU AI Act states that risk means "the combination of the probability of an occurrence of harm and the severity of that harm."

68 The EU AI Act makes a distinction (inter alia) between the provider (the organisation that "develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge" - Article 3(3)) and the deployer (the organisation "using an AI system under its authority" - Article 3(4)).

69 For an overview of the CEN-CLC/JTC 21 work programme set up by the European Committee for Standardisation (CEN) and the European Committee for Electrotechnical Standardisation (CENELEC), visit the CEN website at https://standards.cencenelec.eu/dyn/www/f?p=205:22:0::::FSP_ORG_ID,FSP_LANG_ID:2916257,25&cs=1827B89DA69577BF3631EE2B6070F207D.

70 Regulation (EU) 2024/1689 of 13 June 2024, laying down harmonised rules on artificial intelligence, Art. 6(3).

71 On this particular point, the EU AI Act refers to the GDPR, which provides the following definition (Article 4(4)): "profiling means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements." As the CNIL points out: "Profiling processing involves building an individualised profile of a specific person and aims to evaluate certain personal aspects for the purpose of judging or reaching conclusions about that person (...). Special care must be taken with any profiling processing operations since, by their very nature, they raise significant risks for the rights and freedoms of natural persons." The CNIL gives the following example: "Profiling underlies the insurance sector, particularly when it involves evaluating the characteristics of the insurance risk with the aim of determining the frequency, average cost, maximum cost of a potential claim, pricing and the insurability of the risk. Therefore, it is only a means and not an end": CNIL, Data subjects' rights and profiling: specific characteristics of the insurance sector, at https://www.cnil.fr/fr, published on 16 July 2021.

72 CNIL, Data subjects' rights and profiling: specific characteristics of the insurance sector, at https://www.cnil.fr/fr, published on 16 July 2021.

73 See the earlier discussion about the CJEU ruling on the concept of decision (CJEU, 7 Dec. 2023, no. C-634/21).

[74] Table of the local authority's main algorithmic processing operations that lead to an individual administrative decision, 25 March 2024, published at https://www.valdoise.fr

[75] In pursuance of the last paragraph of Article 47 of the French Data Protection Act: "No decision for which the authorities rule on an administrative appeal stipulated in Part I of Book IV of the Code of Relations between the Public and the Administration may be taken solely on the basis of automated personal data processing."

[76] Article 29 Data Protection Working Party, Guidelines on automated individual decision-making and profiling for the purposes of Regulation (EU) 2016/679, 6 February 2018, p. 23.

[77] Conseil d'État, Annual study 2014, Digital technology and fundamental rights, La documentation française, 2014.

[78] CNIL, deliberation no. 2016-180, 16 June 2016.

[79] CNIL, deliberation no. 2023-058, 8 June 2023.

[80] Constitutional Council, 27 Dec. 2019, no. 2019-796 DC.

[81] Article 29 Data Protection Working Party, op. cit. 23.

[82] See in particular Liane Huttner, The algorithmic decision, private law study into the relationship between humans and machines, Dalloz, 2024.

[83] See in particular Winston Maxwell, Human control of algorithmic systems - a critical look at the need for "humans in the loop", original dissertation for accreditation to direct research, law, Paris 1 Panthéon-Sorbonne University, 2022.

[84] Council of Europe, Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT), Study [DGI(2019)05] of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework, 2019, p. 68; Madeleine Clare Elish, Letting Autopilots Off the Hook: Why do we blame humans when automation fails?, at https://slate.com/, published on 16 June 2016.

[85] Education Code, Art. D. 612-1-13.

[86] This is confirmed in the guidelines from the Ministry of National Education and Youth, the Ministry of Sports, and the Ministry of Higher Education and Research, Use of the Parcoursup decision support module: obligations of the GDPR, 2021, p. 3. "Insofar as the decision is not fully automated, the requirements of the GDPR relating to algorithmic processing, which implies informing data subjects about the logic involved in the algorithm as well as the significance and the envisaged consequences of that algorithm, do not apply to these processing operations."

[87] Constitutional Council, 3 April 2020, no. 2020-824 QPC, §14.

[88] Liane Huttner, op. cit.

[89] Anne Debet, "Parcoursup: the Constitutional Council enshrines a constitutional right for access to administrative documents", Communication - Commerce électronique, no. 9, 2020, comm. 65.

[90] See Jacques Grosperrin, "Local algorithms in Parcoursup - Communication", in. Reports from the Committee on Culture, Education and Communication [Senate], 2019: on Paris Descartes University: "Once pre-ranking had been carried out, human operations began: the Committee looked at all the applications from candidates in the sector - around 600 applications - and the first 100 from outside the sector (...)."

[91] Société informatique de France (SIF), College leavers: the Parcoursup obstacle course, 2024.

[92] French Regulation of 17 July 2017, on the creation of an automated personal data processing system called "Affelnet lycée" by the Ministry of National Education, NOR: MENE1719988A.

[93] Defender of Rights, decision no. 2023-140, 26 June 2023.

[94] Liane Huttner, op. cit.

[95] In this respect, see Nicolas Spatola, Human oversight is still essential for interpreting AI results and taking informed decisions, at https://acteurspublics.fr/, published on 17 July 2024.

[96] See Conseil d'État, Artificial intelligence and public action: building trust, serving performance, 2022.

[97] See Conseil d'État, Artificial intelligence and public action: building trust, serving performance, 2022.

[98] Conseil d'État, Annual study 2014, Digital technology and fundamental rights, La documentation française, 2014. 23.

[99] Winston Maxwell, op. cit.

[100] Applicability of Article 22 of the GDPR, included in Article 47 of the French Data Protection Act for processing operations subject to the GDPR.

[101] ENA, Promotion 2018-2019 MOLIERE, op. cit.

[102] In the case of Affelnet, the scale sheet showing only a value of "0" in the category relating to school assessments would have been checked by taking account of the pupil's school grades.

[103] See in particular Winston Maxwell, op. cit.

[104] Council of Europe, Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT), op. cit.

[105] Winston Maxwell, op.cit.

[106] CRPA, art. R. 311-3-1-2.

[107] Liane Huttner, op. cit.

[108] Jean-Marc Sauvé, Transparency, the values of public action and the general interest, opening speech at the conference entitled "Culture of secrecy versus unlimited transparency: what balance should be struck to guarantee the general interest?", Transparence International France, National Assembly, 2011.

[109] Cf. Article 41 of the Charter of fundamental rights of the European Union, 2000/C 364/01, 2000, which enshrines the "right to good administration": "1. Every person has the right to have his or her affairs handled impartially, fairly and within a reasonable time by the institutions and bodies of the Union. 2. This right includes: a) e right of every person to be heard, before

any individual measure which would affect him or her adversely is taken; b) e right of every person to have access to his or her file, while respecting the legitimate interests of confidentiality and of professional and business secrecy; c) e obligation of the administration to give reasons for its decisions. [...]"

[110] Article 47(2) of the French Data Protection Act stipulates that the decisions concerned must include this explicit information, otherwise "they will be deemed invalid".

[111] For a critical approach to the concept of transparency, see for example: Louis Vuarin, Véronique Steyer. The transparency requirement: a double-edged regulatory lever for organisations, Bulletin of the French Association for Artificial Intelligence, 2023, pp.26-35.

[112] Jean-François Kerléo, Transparency has morphed into a public policy, at https://acteurspublics.fr/, published on 22 April 2024.

[113] According to Article L. 211-5 of the CRPA: "The reasons required by this chapter must be stated in writing and include a statement of the legal and factual considerations on which the decision is based."

[114] Lucie Cluzel-Métayer, *op. cit.*

[115] Félicien Vallet, Interview with Clément Henin, Daniel Le Métayer: "Explain how algorithms work in terms that can be understood by lay users", at https://linc.cnil.fr/, published on 6 January 2021.

[116] Using the example of a decision support system for processing bank loan applications, Clément Hénin and Daniel Le Métayer indicate that one explanation could be: "your loan application has been refused, because your debt ratio exceeds one third of your income". This information explains why the application has been refused, but does not offer any justification. The justification could be: "your loan application has been refused, because the law prohibits banks from granting loans that lead to a debt ratio in excess of one third of your income," *Ibid*.

[117] See in particular the considerations and actions undertaken by the "Le mouton numérique" collective and the "La Quadrature du Net" association.

[118] In this regard, see Simon Chignard, The essential need for transparency in public algorithms, at https://dataanalyticspost.com/, published on 12 June 2019.

[119] Soizic Pénicaud, Simon Chignard, Applying the open government principle to public algorithms, Webinar, at https://www.modernisation.gouv.fr/ [Dinum], recorded on 31 May 2021.

[120] In this respect, see Defender of Rights, CNIL, Report - Algorithms: preventing automated discrimination, 2020.

[121] In this respect, see Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, *op. cit.* p. 68.

[122] See Cada, "Interview with Fabien Tarissan", in Activity report 2022-2023, 2024, pp. 44-47.

[123] Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, op. cit, p. 7. - The example given in this report relates to Cnaf, the national family allowance fund which, in its response to the "Changer de cap"

collective, stated that it is required to "implement a very large number of regulatory reforms within an extremely short timeframe, which is putting a strain on [its] information system."

[124] See Article 71-1 of the Constitution of 4 October 1958 and Article 4, 1° of the Framework Act no. 2011-333 of 29 March 2011 relating to the Defender of Rights.

[125] Liane Huttner, *op. cit.*

[126] Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, op. cit., p. 9. - Obfuscation or masking can be defined as "a series of software writing practices that are contrary to good maintenance and production practices, which limit the understanding of the software's behaviour from reading the source code without, however, changing the software's behaviour during execution."

[127] Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, *op. cit.*

[128] Regulation (EU) 2024/1689 of 13 June 2024, laying down harmonised rules on artificial intelligence, whereas 59.

[129] CNIL, Explainability (AI), at https://www.cnil.fr/.

[130] Nicolas Maudet, Grégory Bonnet, Gaël Lejeune et al., AI & explainability, Bulletin of the French Association for Artificial Intelligence, no. 116, 2022.

[131] In this respect, see the project launched in the summer of 2024 by the CNIL and its digital innovation laboratory (Linc) on a database containing around 12,000 publications: Romain Pialat, Launch of a project on explainability in the field of artificial intelligence, at https://linc.cnil.fr/, published on 1 August 2024.

[132] Félicien Vallet, *op. cit.*

[133] *Ibid.*

[134] "This field still suffers significantly from what researcher Tim Miller and his colleagues have called the "beware of inmates running the asylum" syndrome, in this case AI experts explaining their own productions", *ibid.*

[135] In this respect, reference can be made to the work revealing how explainability has been massively replaced by performance indicators. See Louis Vuarin, Véronique Steyer, "The principle of explainable AI and its application in organisations", in the Réseaux - Communication - Technologie - Société review, no. 240 (4), pp. 179-210.

[136] CNIL, GDPR compliance: how to inform individuals and ensure transparency, at https://www.cnil.fr/, published on 29 July 2019.

[137] Cf. Cada, Cada's role, at https://www.cada.fr/.

[138] Cf. Guidelines on automated individual decision-making and profiling for the purposes of Regulation (EU) 2016/679, which state that even if the automated decision-making and profiling does not meet the definition in Article 22 of the GDPR, it is nevertheless "good practice to provide the information" (p. 28).

[139] In accordance with Articles L. 311-3-1 and R. 311-3-1-1 of the CRPA.

[140] Secrets listed in Article L. 311-5, 2° of the CRPA.

141 In accordance with Article R. 311-3-1-2 of the CRPA.

142 In other words, the State, as specified in Article L. 300-2 of the CRPA, regional or local authorities, or other public or private legal entities entrusted with a public service mission.

143 In accordance with Article D. 312-1-4 of the CRPA.

144 Secrets listed in Article L. 311-5, 2° of the CRPA.

145 Cada, CNIL, Practical guide to the online publication and reuse of public data (open data), 2016, p. 26: "Before an algorithm can be implemented by a computer, it must be expressed in machine language and transcribed into a program (a text comprising written commands, also known as 'source code')."

146 CRPA, Art. L. 300-2.

147 A request for access to administrative documents, also known as a "Cada request", enables any person to submit an oral or written request for access to an administrative document. The following page on the service-public website explains the procedure: https://www.service-public.fr/particuliers/vosdroits/F2467.

148 CRPA, Art. L. 311-9.

149 As specified in Article L. 312-1-1 of the CRPA, these authorities must publish online the documents that they disclose, subject to protected secrets.

150 CRPA, Art. D. 312-1-1-1.

151 Regulation (EU) 2016/679 of 27 April 2016, GDPR, Art. 15.

152 Regulation (EU) 2016/679 of 27 April 2016, GDPR, Art. 30.

153 Constitutional Council, 12 June 2018, no. 2018-765 DC.

154 CRPA, Art. R. 311-3-1-2.

155 Source: CNIL, Parameter (AI), at https://www.cnil.fr/.

156 Winston Maxwell, op. cit.

157 Article 86 of the EU AI Act states that this right applies only insofar as it is not provided for elsewhere in Union law and that this right applies unless exceptions are provided for in EU or national law.

158 In accordance with Article L. 311-5, 2° of the CRPA.

159 Note that for some authorities, such as the Defender of Rights, only the secrecy of national defence, State security and foreign policy may be binding (Article 20 of the Framework Act no. 2011-333 of 29 March 2011 - on this particular subject, refer to: CNIL, Compendium of procedures for authorised third parties, 2020).

160 See Cada, Activity report 2022-2023, 2024, p. 42, regarding its opinion no. 20213847 of 13 January 2022.

161 Cada, opinion no. 20226179 of 15 December 2022.

162 According to the decision of the Constitutional Council of 3 April 2020, no. 2020-834 QPC: "Application of Article L. 311-3-1 of the aforementioned code, which would require the government agency, subject to secrets protected by law, to disclose upon request to the applicant the rules defining the algorithmic processing operation and the main characteristics of its implementation, is excluded. Application of Article L. 312-1-3 of the same code, which would require the government agency, subject to the same proviso, to publish online the rules defining the main algorithmic processes used in the performance of its duties, is also excluded."

163 CRPA, Art. L. 311-6.

164 Cada, op. cit., p. 43: regarding its opinion no. 20230314 of 30 March 2023.

165 CE, 8 November 2017, no. 375704.

166 In accordance with Article L. 300-2 of the CRPA, this concerns the State, regional or local authorities, or other public or private legal entities entrusted with a public service mission, with over 50 FTEs (Article D. 312-1-4 of the CRPA).

167 In its aforementioned report on AI, the Conseil d'État states that "there are doubts surrounding this obligation's effectiveness in practice" (p. 120); Lucie Cluzel-Métayer, op cit, indicates that "few public service agencies have initiated this delicate task of mapping their processing operations"; Emile Marzolf, Algorithms: the Gordian Knot of administrative transparency, at https://acteurspublics.fr/, published on 17 April 2024: "In practice, algorithmic transparency is very rarely implemented."

168 Emile Marzolf, The authorities in the Ille-et-Vilaine region lift the curtain on their algorithms, at https://acteurspublics.fr/, published on 24 July 2024.

169 In this respect, it is worth mentioning: France Travail, Algorithms: everything you wanted to know about algorithms published by France Travail, at https://www.francetravail.fr/; Ille & Vilaine, List of algorithmic processing operations, at https://data.ille-et-vilaine.fr/; Val-d'Oise region, Table of the local authority's main algorithmic processing operations that lead to an individual administrative decision, 2024.

170 Visit the website of the Ouvre-boîte association, whose aim is to obtain the effective publication of administrative documents: https://ouvre-boite.org/.

171 Visit the website of the ODAP public algorithm observatory project, which creates and gathers information about the algorithms used by the French authorities: https://odap.fr/.

172 See Emile Marzolf, The parties responsible for access to administrative documents, a little known cog in the information wheel, at https://acteurspublics.fr/, published on 17 April 2024.

173 The Guide is available online at https://guides.etalab.gouv.fr/algorithmes/mention/#dans-quels-cas-l-obligation-de-mention-explicite-s-applique-t-elle.

174 In this respect, Dinum's roadmap makes no mention of the transparency of public algorithms: Dinum, A digital strategy for effective public action, DINUM Roadmap, 2023.

175 CRPA, Art. R. 311-3-1-1 and R. 311-3-1-2.

176 For example, this was the case in the facts that led to the Defender of Rights' decision no. 2023-140 of 26 June 2023.

177 In particular, refer to the page on the request process created by the Ouvre-boîte association and the Ma dada initiative of the OpenKnowledge France

association.

[178] https://www.ipsos.com/fr-fr/51-des-francais-ne-se-declarent-pas-particulierement-effrayes-par-lintelligence-artificielle

[179] Markus Krajewski, "Source Code Criticism: On Programming as a Cultural Technique and its Judicial Linkages", in. Journal of Cross-disciplinary Research in Computational Law 1.3, 2023.

[180] Jenna Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms", in. Big Data & Society 3.1, 2016.

[181] See in this regard: Dinum, Free software and digital commons action plan, at https://www.numerique.gouv.fr/, published on 10 November 2021.

[182] In this regard, see: Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, op. cit: "Open source publication is a necessary, but insufficient condition for the authorities' transparency towards the public"; "not all decisions taken on the basis of automated processing are explained, and when the responsible party provides explanations, they are characterised by their brief or even hermetic nature"; "making documents available on the Internet is far from resolving the issue".

[183] Winston Maxwell, *op. cit.*

[184] Constitutional Council, 12 June 2018, no. 2018-765 DC.

[185] Proposal made in particular by Judith Rochfeld, "Personal data - Right not to be subject to a decision based solely on automated processing", in. IP/IT and Communication, 2020, p. 23.

[186] See Conseil d'État, Artificial intelligence and public action: building trust, serving performance, 2022.

[187] Denis Merigoux, Marie Alauzen, Louis Gesbert et al., INRIA, *op. cit.*

[188] Based on the Catala project developed by INRIA (programming methodology that links the system's source code to the texts on which its automated decision is based).

[189] Tool developed by Clément Hénin and Daniel Le Métayer. For example, see Félicien Vallet*, op. cit.*

[190] These three avenues are suggested by Winston Maxwell, *op. cit.*

[191] Cada, "Interview with Fabien Tarissan", in Activity report 2022-2023, 2024, pp. 44-47.

[192] In this regard, see Lucas Boncourt, The challenge of explainable public algorithms, at https://www.banquedesterritoires.fr/, published on 17 September 2019.

[193] This comes at a time when some stakeholders have reported that the obligation to publish the rules defining the main algorithmic processes that the authorities use to perform their duties when they form the basis of individual decisions has remained a "dead letter" since 2016. In this respect, see: Emile Marzolf, The authorities in the Ille-et-Vilaine region lift the curtain on their algorithms, at https://acteurspublics.fr/, published on 24 July 2024.

[194] Jean-François Kerléo, Transparency has morphed into a public policy, at https://acteurspublics.fr/, published on 22 April 2024.

[195] See La Quadrature du Net, Algorithmic CCTV: surveillance builds its empire, at https://www.laquadrature.net/.

Defender of Rights - TSA 90716 - 75334 Paris Cedex 07 - 09 69 39 00 00

defenseurdesdroits.fr

f **X** **in** **O**

D

Défenseurdes**droits**
RÉPUBLIQUE FRANÇAISE