“With great power comes great responsibility”: keeping public sector algorithms accountable

Simon Chignard, Soizic Penicaud
Etalab, French Prime Minister task force for open data & open government
simon.chignard@data.gouv.fr
soizic.penicaud@data.gouv.fr

Abstract

Computer algorithms are increasingly used in public sector. In France, heart transplants allocation, calculation of taxes and social benefits, and matching between schools and students are some examples of algorithm-mediated administrative processes. Although nothing new for the public sector, public sector algorithms (PSA) offer new forms of public policy and thereby enable automation and personalisation of individual interactions and decisions. But these instruments, as powerful as they are, present new challenges for society and individuals especially in terms of transparency and accountability. On this topic, the French context is shaped on one hand by the opportunity of a new regulatory framework for PSA accountability and, on the other, by the new challenges posed by machine learning algorithms. In this paper, we explain how Etalab, the French government’s task force for open data and data policy, is working with public administrations to keep PSA accountable.

1 - From automation of human tasks to machine learning: a short history of public sector algorithms

Algorithms are nothing new for the public service. In the 1970’s, the emergence of the computing industry allowed the State to automate some of its administrative processes, starting with the large-scale and low-variability processes. Tax calculation and social benefits management were among the first applications of computed algorithms; both tax calculations and social benefits management affect a lot of households (approximately 37 million for income taxes) and the rules of calculation easily translate into software before being operated by computers. In this first generation of public sector algorithms computers replace calculators and accountants to achieve the same tasks but at a completely different scale, speed, and accuracy.

The second generation started with the use of matching algorithms, mainly for human resources management. The Ministry of Education developed a system to manage their workforce, that is, educators and teachers applying for a new position in a different school or region. Since the early 2000’s, matching algorithms have been used for student allocation such as Affelnet, Admission Post-Bac, and Parcoursup. However, these systems have been criticized for their complexity and lack of transparency. Again, this process affects a lot of individuals but the use of algorithms in these cases are not a strict replacement of humans. Computer algorithms made it possible to accomplish tasks like matching people to positions at a national scale, something that was unachievable before. For example, Parcoursup, the higher education admission system processes 7 million of admission requests from 850, 000 students each year; it is quite impossible for companies such as Parcoursup to deal with such a great number of requests without a computerized system. As such, algorithms are now used not only to automatize but also to augment the delivery of public services.

The third generation of public sector algorithms is linked to the emergence of machine learning (ML) algorithms which represent a major shift for public policy. In the first two periods, computers were used to apply a set of rules that were sometimes complex, but always predefined. Instead, ML algorithms derive rules from observations and learning from large datasets. For example, La Bonne Boîte (a service designed by Pole Emploi, the French employment agency) can predict the hiring potential in the next 6 months of *any* company for *any* job in a specific city or region. The algorithm is learning from around 10 million hiring declarations. ML algorithms are also used for tax fraud detection or prediction of companies that present a risk of going bankrupt (see the Signaux Faibles project). Machine learning algorithms change the very nature of public action, from the application of pre-defined rules to the construction of new rules through observation.

i. Automation, augmentation and transformation

	1st generation	*2nd generation*	*3rd generation*
Type of algorithms	Calculations	Matching algorithms	Machine learning algorithms
Main effect on public policy	Automation	Augmentation	Transformation
Period	1970-now	1990-now	2015-now
Type of change	Scale, speed & accuracy	Scope	Nature
Examples	Impots.gouv.fr (tax calculation)	Parcoursup	La Bonne Boîte

ii. Benefits of using algorithms

From an administration point of view, there are four main benefits of using computer algorithms:

1. *speed*: computers are able to calculate millions of individual situations in a very short time;
2. *cost*: compared to human agents, computers can do the same basic task at a tiny fraction of the budget;
3. *accuracy*: computers can deal with complex situations and provide a reliable result;
4. *new capabilities or scope*: some applications were unthinkable or impossible to achieve before the introduction of computers.

Reach is a less known or discussed, fifth, additional benefit. Automated systems enable public administration to extend the reach of public action to every part of the territory.

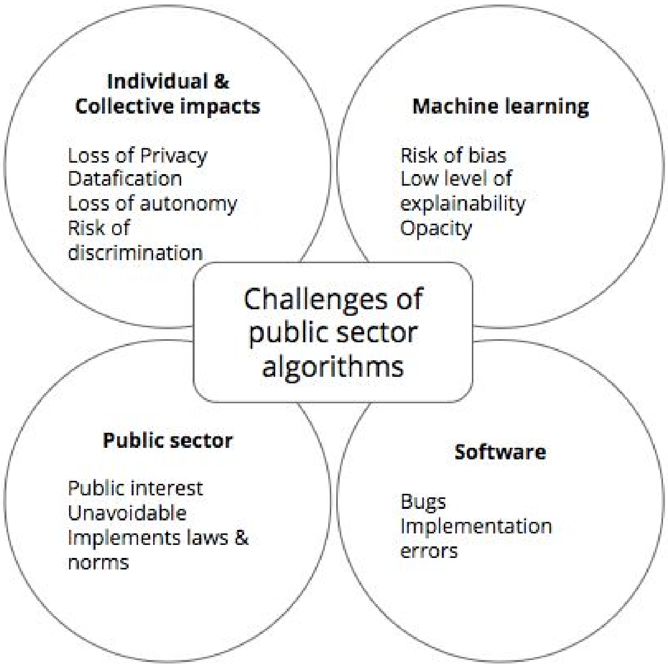
For example, the use of automatic speed camera linked to a central system is a way of implementing a public policy (speed reduction) at a national scale from a single point of decision. Before the implementation of this centralised system each local authority was able to define its own plan for vehicle speed control (place and time of control, degree of severity of the sanction). As such, centralised systems can be seen as a way of expanding the reach of public administration.

Algorithms are now part of public action. An additional change has been introduced by artificial intelligence technologies in national and local governments. This entails changes for public action, not in terms of speed or scale but also in its nature. Defining rules is at the heart of the policy-making process and is usually preliminary to the creation of the tool. Artificial intelligence blurs the frontier between the public policy and the tool used to implement it. It is no surprise that the Constitutional Council, responsible for the review of the constitutionality of legislation, stated that algorithms that are capable of revising the rules that they apply, without human control and validation may not be used as the sole basis for an individual administrative decision.

2 – The challenges of algorithms for individuals and society

Public sector algorithms imply a series of challenges related to their algorithmic nature (software, machine learning challenges), to their field of application (public sector and public policy) to the people and organizations impacted (on their individual and collective rights).

As a piece of software, algorithms are subject to conception and implementation errors such as bugs or biased datasets. Machine learning systems are particularly prone to bias inherited from the learning datasets. Opacity is another challenge of algorithms, be it from the technical complexity or from a deliberative action to obfuscate the algorithm itself. Low-level of explainability and potential loss of autonomy for the people are well-documented risk of algorithms.



i. Specific challenges of public sector algorithms

All these challenges are not specific to the public sector, which does not mean that public and private sector algorithms are the same subject. Algorithms used by administrations are distinguishable in three ways worth noting.

First, they must be used in the **public interest** and not a *particular or private* interest. One can question the impact of YouTube's recommendation algorithm on the diffusion of harmful content. That said, it is generally accepted that, as a private company, YouTube is pursuing its own interest and not the public interest. Second, in many cases, public sector algorithms tend to **implement legal rules**. Tax calculations follow a list of rules defined by the general tax law and adopted by the Parliament. As such, some PSA are the last link from the political will to tangible effects on individuals.

Finally, public sector algorithms tend to be unavoidable: citizens do not have the option to use a different algorithm and sometimes are not presented with the choice to opt out of using an algorithm at all. For example, a French patient in need of a heart transplant has no choice but to accept the rules and their algorithmic implementation by the Biomedicine agency. Similarly, the only way to get access to most of the higher education institutions is to go through Parcoursup, the allocation system for students, which relies on an algorithm.

In 1789, the Declaration of the Rights of Man and of the Citizen (article XV) stated that "*Society has the right to require of every public agent an account of their administration*". This principle, enacted well before the invention of computers, still stands: if administrations use algorithms as tools of administration and government, then they – and their algorithms – should be held accountable.

3 - A framework for public sector algorithms accountability

In this section, we present a general framework to help public sector agencies to achieve accountability. This framework is composed of a series of guiding principles and a legal basis, be it from GDPR (european) or the Digital Republic bill (national).

i. Guiding principles for keeping algorithms accountable

Public sector algorithms – and first the agencies that use them – should be held accountable. To help agencies practice accountability we devised of list of six guiding principles that each public agency could follow, independently of the purpose of the algorithm:

1. *acknowledgment*: agencies should mention when an algorithm is used;
2. *general explanation*: agencies should provide a clear and understandable explanation of how algorithm works;
3. *local (individual) explanation*: agencies ought to provide a personalised explanation of a specific result or decision;
4. *justification*: agencies should justify why an algorithm is used and reasons for choosing a particular algorithm;
5. *publication*: agencies should publish the source code and documentation as well as note whether or not the algorithm was built by a third party;
6. *allow for contestation*: agencies should provide ways of discussing algorithmic processes and appeal them.

ii. A new legal basis to implement these principles for individual decisions

Several texts of national and European law have recently come to regulate the use of algorithms in France, particularly when they are used to make decisions that affect individuals. The Digital Republic Bill (2016) introduced the principle of transparency of public algorithms. In practice, this means offering individuals (natural and legal persons alike) new rights. The text lays down three obligations for administrations:

- *explicit mention*: i.e. administrations are obligated to inform interested parties that an algorithm is being used and what the person's legal rights are;
- *general information*: administrations must publish the operating principles of the main processing operations when these operations are the basis of individual administrative decisions;
- *Individual information*: administrations should provide the individual concerned with a detailed set of information about the algorithm, its functioning and the data processed for the individual's specific case.

This right to individual information, if claimed by individuals and applied by administrations, can lay the foundations for a more peaceful relationship between an increasingly algorithmic administration and a society that is increasingly concerned by these tools. Compared to GDPR, the obligations introduced by the Digital Republic Bill are broader, as they cover both automated decisions and the equally numerous cases where algorithms are only decision support tools.

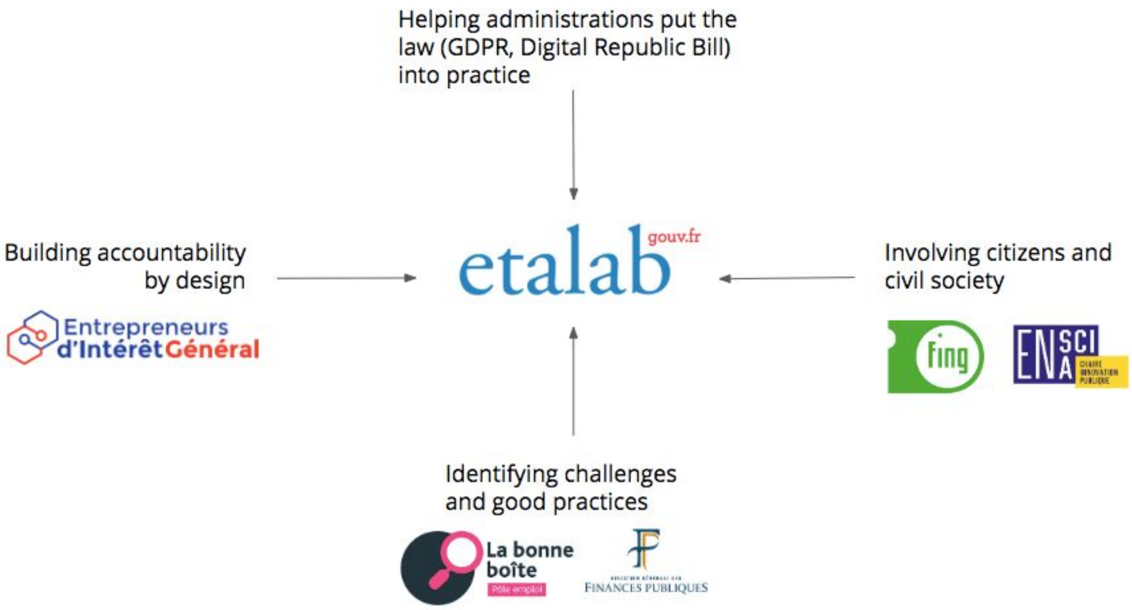
4 - How Etalab is working towards PSA accountability

With the introduction of a new legal framework for algorithmic accountability and transparency, public agencies need to be accompanied on making existing and future algorithms compliant with the new obligations. This also gives citizens access to new rights, such as an extended right to information.

Through the National Action Plan (2018-2020) for the Open Government Partnership, France has committed to reinforcing “the transparency of public sector algorithms and source codes”. Etalab, as the government’s task force for open data and data policy, oversees the work on this commitment which lies at the crossroads between open data, open source, and open government issues.

Our approach is multifaceted:

- *top-down*: from the general principles and legal obligations to their concrete implementations;
- *bottom-up*: deriving challenges and good practices from specific case studies;
- *inspired by inputs*: from external partners and from alternative methods;
- *next to*: supporting teams developing new algorithms.



i. Putting the law into practice

To answer to the agencies’ primary need we published a how-to guide on algorithms for public sector agencies. The guide, available online (in French) – open to contributions on GitHub – gives a brief overview of issues surrounding PSA and puts the European and national legal obligations into layman’s terms to make them more accessible to public servants.

ii. Identifying challenges and good practices

To identify challenges and good practices already in place, we adopted a case-based approach, working with voluntary public sector agencies on existing algorithms. Three different methodologies have been used, each one focusing on a particular object:

a) observing people at work, and their interactions with algorithmic systems;
b) observing code being published, and interviewing developers about their choices;

c) observing legal and practical documentation about administrative procedures using algorithms.

(a) Case #1: Allocating daycare places at a municipal level

We carried out observations as close as possible to the administrative work of the actors, inspired by the concept of *street-level bureaucracy*. In this particular case, algorithms are used to classify requests according to priority criteria (such as occupation of parents, incomes, and special needs). The final decision of accepting or denying a daycare place to a family is made by a commission of elected, professionals and parent representatives based on the classification made by the algorithm. The way these criteria are implemented plays an important role in shaping the discussion of the commission. Participants in the procedure are not always aware of the real impacts of using a particular method of prioritizing applications since the algorithms are woven in the administrative procedure itself and tend to be undetectable. Making them visible is the first step to keeping the administration accountable.

(b) Case #2: Helping the unemployed target potential employers

Pôle Emploi is the French governmental agency which registers unemployed people, helps them find jobs, and provides them with financial aid. La Bonne Boîte focuses on unsolicited job applications by providing job seekers a machine learning-based prediction about the hiring potential in the next six months for any company for each job in a specific location (city, region, etc.). Since La Bonne Boîte publish some part of their source code, we decided to start by analyzing this code. This *whitebox approach* was complemented by a series of interviews with the development team: why do they choose to implement this particular scoring algorithm? What kind of explanations do they provide to their users (and not only to experts) to assist them in reading a source code? This case illustrates the value of mixing approaches and methods to understand an algorithm. Algorithms generate points of different and sometimes conflicting interests: from users, from developers, as well as from policymakers.

```
from labonneboite.common import mapping as mapping_util
from labonneboite.conf import settings

# scores between 0 and 100
SCORE_FOR_ROME_MINIMUM = 20
SCORE_ALTERNANCE_FOR_ROME_MINIMUM = 10

# stars between 0.0 and 5.0
STARS_MINIMUM = 2.5
STARS_MAXIMUM = 5.0

# ############### WARNING about matching scores vs hirings ################
# Methods scoring_util.get_hirings_from_score
# and scoring_util.get_score_from_hirings
# rely on special coefficients SCORE_50_HIRINGS, SCORE_60_HIRINGS etc..
# which values in github repository are *fake* and used for dev and test only.
#
# The real values are confidential, stored outside of github repo,
# and only used in staging and production.
#
# This is designed so that you *CANNOT* guess the hirings based
# on the score you see in production.
# ########################################################################
```

Extract from the source code La Bonne Boîte

(c) Case #3: Allocating heart transplants at a national level

In France, more than 5,000 people are currently living with a transplanted heart. The procedure for allocating heart transplants is managed by the Biomedicine Agency. It must be fair (given the profile of patients and the disparity of regional needs), efficient (providing the community with the maximum benefit expected from this treatment), transparent (and based on the most objective criteria possible), and realistic (taking into account the logistical constraints).

For this particular case, we focused on published documents (whether legal or implementation). The Biomedicine Agency publishes the general scoring procedure, as well as its details. Each time a heart transplant is available the system computes a score for every patient on the waiting list. This score is based on different criteria (morphology, bio-compatibility, level of priority and urgency, etc.). Different functions are used including the following one based on the matching of age between donor and receiver. In each case, the objective is to increase the life duration of the receiver. The lower the age difference between donor and receiver, the higher the matching score, this to avoid that an older heart be transplanted into a younger body.

1.2.3 ETAPE 3 : CALCUL DU SCORE CARDIAQUE COMPOSITE PONDERE (SCORE CCP)

Le calcul du Score Composite Pondéré résulte de l'application au Score Cardiaque Composite Brut (Score CCB) d'un ensemble de filtres et de fonctions d'appariement donneur – receveur. Ces filtres s'appliquent lors de la proposition du greffon cardiaque.

1.2.3.1 Fonction d'appariement en âge entre receveur et donneur

La différence d'âge entre le receveur et le donneur est prise en compte en appliquant un facteur de pondération aux nombre de points donnés par les Composantes Adulte *(voir : ANNEXES 3.4.1)* :

Si le donneur est plus jeune que le receveur :

- ✓ différence d'âge **≤ 15 ans :** → 100% des points
- ✓ différence d'âge ≥ **15 et ≤ 40 ans :** → un pourcentage des points décroissant à partir de 100% jusqu'à 0%
- ✓ différence d'âge **> 40 ans :** → 0% points

Si le donneur est plus âgé que le receveur :

- ✓ différence d'âge **≤ 40 ans :** → 100% des points
- ✓ différence d'âge ≥ **40 et ≤ 65 ans :** → un pourcentage des points décroissant à partir de 100% jusqu'à 0%
- ✓ différence d'âge **> 65 ans :** → 0% points

3.4.1 Fonction d'appariement en âge entre donneur et receveur

```
ΔAgeRD = (AGER-AGED)
F0_DifAge =     Si ΔAgeRD <0  Alors
                              (ΔAgeRD +40)/25
                Sinon
                      1-(ΔAgeRD -15)/25
                Fin Si
F1_DifAge = Si (AGER>=18) Alors
                   min(1;max(0; F0_DifAge))
                Sinon
                       1
                Fin Si
```

The function that matches the age between donor and receiver, and its explanation "in plain language". Both images taken from the "Guide du Score Coeur" (Heart Score Guide)

This case presents a way for an agency to make its technical choices explicit, and link them to the public policy choices they derive from.

All in all, this case-based approach paints a more precise picture of what public sector algorithms can look like. This proves to be very useful because "algorithms" are a blurry category and public agencies sometimes struggle to identify if they are currently using algorithms. Illustrations of existing algorithms can help agencies pinpoint that their Excel spreadsheet is, indeed, sometimes an algorithm.

iii. Building accountability by design

In addition to working with existing use cases, we explore accountability by design by working in close collaboration with different organizations and actors.

a. Supporting innovation programs currently developing algorithms

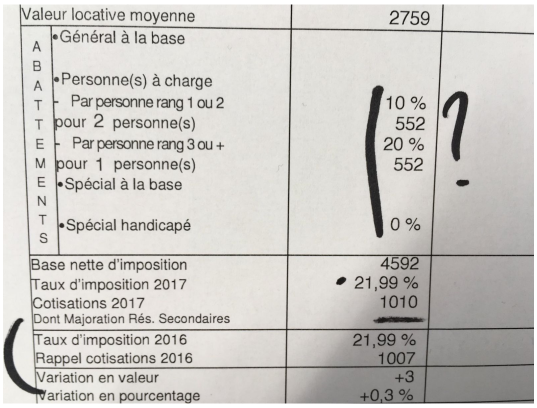
As the department for data policy, Etalab encourages the development of datascience into government through different innovation programs developing rule-based and machine learning algorithms, such as digital transformation program Entrepreneurs d'Intérêt Général and a call for projects around artificial intelligence.

We make sure that PSA accountability issues are addressed from the beginning of these programs by organizing workshops and discussions with the development teams (public servants, data scientists, and developers). We work on different topics such as impact, explanation, data bias or symmetry of information. Working with these teams allows us to test tools such as the San Francisco's and Harvard's Ethics & AI Framework. Somewhat more importantly, these workshops open a space where development teams can reflect together on the tools they create, and the impact they may have on users and citizens.

b. Involving users and civil society in the issue

To experiment with involving citizens in making PSA accountable we conducted two participatory workshops on the French housing tax, with the help of PhD candidate Loup Cellard (Warwick

University). We used the analogy of an algorithm as a cooking recipe, with ingredients (data), step-by-step preparation methods (instructions) and a final result (amount to be paid).



Valeur locative moyenne	2759
ABATTEMENTS •Général à la base	
•Personne(s) à charge	
- Par personne rang 1 ou 2 pour 2 personne(s)	10 % 552
- Par personne rang 3 ou + pour 1 personne(s)	20 % 552
•Spécial à la base	
•Spécial handicapé	0 %
Base nette d'imposition	4592
Taux d'imposition 2017	21,99 %
Cotisations 2017	1010
Dont Majoration Rés. Secondaires	
Taux d'imposition 2016	21,99 %
Rappel cotisations 2016	1007
Variation en valeur	+3
Variation en pourcentage	+0,3 %

A participatory workshop on housing tax (Paris, 2018)

Getting the participants' input on how to better explain the tax calculations was important but the workshop was particularly useful for raising awareness around citizens' right to information with regards to public sector algorithms.

c. Fostering dialog between public agencies and research organizations

We work in close collaboration with people involved in these topics outside of government, and researchers in particular, to create links between them and voluntary agencies. This allows us to explore less traditional pathways left untrodden by the public sector.

For example, think tank FING (Fondation Internet nouvelle génération) and design school Ecole Boulle conducted a design sprint with public sector algorithm Signaux Faibles. Together they asked the question: how can design help bridge the information gap between those who use algorithms and those who are impacted by them? Take Signaux Faibles which use machine learning to better detect companies at risk of failure: this information is used by public authorities to offer assistance to entrepreneurs. Being targeted without exactly understanding why can be a stressful process for company owners. With this in mind, design students reworked the tool's interface so that it was not perceived as a threatening tool by the companies identified but, rather, as a self-monitoring instrument. This case illustrates how design and user-centered methods can help reduce the asymmetry of information between those who use algorithms and those who are impacted by them.

Signaux Faibles re-design by Fing and Ecole Boulle (2018)

5 - Perspective: public sector accountability and citizens

So far, our department's missions have been mainly focused on making sure public agencies meet their legal obligations. However, these obligations only make sense if they genuinely enable citizens to exert their rights, and if they result in a juster way to conduct public action.

This raises crucial questions we want to keep exploring: who from outside the public sector should be involved in public sector algorithm accountability and at which step of the process? Secondly, where can and should citizens play a role? How can they help avoid the pitfall of "accountability-washing" and focus the discussion beyond the algorithm and on the public policies at stake?

These questions call for broader reflection at two levels: first, nationally, by mobilizing civil society organizations, the media, human and digital rights NGOs, and research institutions while taking into consideration the local social, political and legal contexts. Secondly, through international collaborations with governments and organizations that have been working on these topics.

To tackle this, we are among other things co-organizing a session at RightsCon Summit in Tunis on June 13, 2019. We hope to hear from many initiatives around the world and we are eager to exchange lessons learned and good practices.

Acknowledgments

The authors would like to thank the Etalab team for their contribution, and more specifically Bastien Guerry and Amélie Banzet for their review and comments, as well as Shani Bans for her editorial help. Our work on public sector algorithms also benefits from the contribution of Loup Cellard (Warwick University), Hubert Guillaud and Thierry Marcou (Fondation Internet nouvelle génération).