REVISTA DA FACULDADE DE DIREITO DA UNIVERSIDADE DE LISBOA

LISBON LAW REVIEW

Número Temático: Vulnerabilidade(s) e Direito

ANO LXII 2021 NÚMERO 1 | TOMO 1

REVISTA DA FACULDADE DE DIREITO
DA UNIVERSIDADE DE LISBOA
Periodicidade Semestral
Vol. LXII (2021) 1

LISBON LAW REVIEW

PROPRIEDADE E SECRETARIADO

Faculdade de Direito da Universidade de Lisboa
Alameda da Universidade – 1649-014 Lisboa – Portugal

EDIÇÃO, EXECUÇÃO GRÁFICA E DISTRIBUIÇÃO
LISBON LAW EDITIONS

Alameda da Universidade – Cidade Universitária – 1649-014 Lisboa – Portugal

ISSN 0870-3116 Depósito Legal n.º 75611/95

Data: Agosto, 2021

ÍNDICE 2021

TOMO 1

TOMO 2

VIDA CIENTÍFICA DA FACULDADE

LIVROS & ARTIGOS

Vulnerability and the Algorithmic Public Administration: administrative principles for a public administration of the future

Vulnerabilidade e Administração Pública Algorítmica: princípios administrativos para uma Administração Pública de futuro

Francisco de Abreu Duarte* | **Rui Tavares Lanceiro****

Abstract: This article considers some of the challenges arising out of the use of information technologies (IT), artificial intelligence, and automated means in public decision-making. It aims at understanding the risks posed by such technologies for citizens and inquires over the future of administrative law and the protection of fundamental rights in such context. The paper begins by show-casing a series of examples of the use of algorithms by public administrations in the areas of education, criminal justice, security and social security. It then proceeds to analyzing whether traditional administrative principles are still suitable to address these new challenges or whether it is necessary to create new rules to safeguard fundamental rights. It concludes with a call for further research, arguing for a procedural view of administrative law. It is ultimately defended that the foundational principles of administrative law are sufficient in principle but must be updated to face the new challenges of the algorithmic age.

Resumo: O presente artigo visa compreender os desafios trazidos pelo uso de tecnologias de informação e decisão-algorítmica pela Administração Pública, tecendo considerações iniciais sobre a adaptabilidade do Direito Administrativo a esta nova realidade. O trabalho começa por demonstrar as vulnerabilidades existentes na relação administrativa algorítmica, ilustrando casos de estudo em que se denotam os potenciais riscos para os direitos fundamentais nas áreas da educação, justiça penal, segurança e segurança social. Através da análise destes casos, é possível questionar acerca da pertinência e atualidade dos princípios de direito administrativo clássicos, assim como da sua potencial adaptabilidade a decisões administrativas algorítmicas. O artigo conclui defendendo que os princípios clássicos de Direito Administrativo mantêm a sua pertinência, mas carecem de uma atualização urgente para fazerem face aos desafios da nova revolução algorítmica.

* PhD candidate European University Institute; LL.M European University Institute; LL.M New York University (NYU); Associated Research Fellow at the Lisbon Centre for Research in Public Law.
** Assistant Professor at the University of Lisbon Law School; Senior Research Fellow at the Lisbon Centre for Research in Public Law.

Keywords: artificial intelligence; algorithmic administration; administrative law; transparency; accountability.

Palavras-chave: inteligência artificial; administração algorítmica; direito administrativo; transparência; responsabilidade.

Introduction

Public Administration is facing a profound change. The "new industrial revolution", the "fourth industrial revolution[1]", the "second machine age[2]", are different expressions that illustrate the process of rapid transformation occurring within contemporary societies. On a global scale, this transformation reaches not only the productive, commercial, and service sectors, but also social structures and public administrations. The perhaps better-called "algorithmic revolution"[3], is bringing about new ways of interaction between citizens and public administrations. Governments around the world are now making use of artificial intelligence (AI) methods to make choices on public policy, ranging from criminal law to education, social scoring, or security. The use of information technologies (IT), artificial intelligence, and automated means in public decision-making and communication with citizens is revolutionizing public administration. New IT tools can improve the efficiency and effectiveness of administrative bodies. The use of artificial intelligence in public decision-making is already happening – at least at the level of the so-called "technical discretion" – reducing the risk of human error and the possibility of liability of the administration. In the same way, the use of automated or standardized forms of communication or customer service are gaining ground. New institutional and financial relationship models between the state and autonomous territorial entities (federal states, regions and local authorities) may also emerge through this way. However, risks and dangers may arise from the use of these new instruments. The human element is essential when the decision-making involves balancing or weighing different interests. Moreover, new problems may arise for

[1] M. Skilton and F. Hovsepian, *The 4th Industrial Revolution: Responding to the Impact of Artificial Intelligence on Business,* Palgrave, (2018).
[2] E. Brynjolfsson and A. McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (2014).
[3] L. Floridi, *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality* (2014).

the respect of fundamental rights of citizens and administrative accountability when a decision is taken through artificial intelligence.

These technologies of "e-government" vary greatly but they all represent a new instrument of public power by public administrations.

The impact of algorithms in the development of administrative relationships is largely uncharted territory. It is important to start discussing the legal frameworks and sets of questions that might guide this topic in the future, in anticipation for everything that might be arriving soon. An "algorithmic administration" – meaning the use of algorithms in administrative decision-making – poses challenges to how we conceive classic administrative relationships, the role of administrative principles such as transparency and due process, and to the attribution of responsibilities and liability. It should make us rethink how existing administrative law doctrine might no longer be fit to address these challenges.

In this short paper, we aim at establishing a mapping exercise of the vulnerabilities that modern citizens can be subject to in the face of such a growing algorithmic public administration. For that effect, the paper is developed along two sections.

Section I starts by providing an overview on the risks of AI technology when used by public administrations. Here we develop on different cases, both judicial and extrajudicial, that show the current challenges that AI technology might bring. Although we do not depart from any preconceived notion on technology as deterministic or neutral[4], the aim of this section is to frame the problem through practical examples. By describing what has happened so far, we provide an overview of the risks, some with proven consequences in many countries. This will guide us to Section II in framing the solutions to those problems.

Section II presents some pathways and a research agenda to face such problems. Here we defend that any algorithmic Administration must necessarily abide by a set of administrative principles which limit its power and protect citizens. We argue that existing administrative principles are flexible enough to encompass this new reality, but they should be subject to an update. In this section, we argue for a procedural approach to the algorithmic age. Although the creation of a detailed sets of substantive norms might seem appealing *prima facie*, the complexities of technological regulation render any such attempt a mere *catching-up* exercise. As technology tends to develop faster than the legislative pace, it is important for scholars to help policymakers in developing general procedural frameworks which can protect fundamental rights in every situation,

[4] See Kranzberg First Law of Technology - Kranzberg, 'Technology and History: 'Kranzberg's Laws", 27 *Technology and Culture* (1986) 544. pp. 545-546.

regardless of the exact content at play. We claim that such rules must be of a procedural nature.

Finally, it is important to note that this paper is just the beginning of a long road ahead. The algorithmic society will be soon upon us, bringing along the wonders of technology that so many find appealing. However, it will also bring a pernicious side, giving rise to many new and unprecedented challenges for fundamental rights. To prepare and build resilience must be the priority.

Section I – Framing the Problem

In this section we intend to present the dangers and pitfalls associated with the use of IT, artificial intelligence, and automated means in public decision-making. In order to do so, we will begin by providing a practical example inspired in a real-life situation.

Imagine, then, the following hypothetical case:

> *Maria is a high-school teacher in Aveiro. She is often praised by her students and students' parents as a fantastic teacher. She has received several awards for best teacher in her school.*
>
> *In 2022, the Ministry of Education decides to start a broad evaluation program for all the high-school teachers in Portugal. The idea is simple: to develop an artificial intelligence system that will rank teachers and allow the Ministry to choose between which teachers to keep and which to let go.*
>
> *The AI system is developed by the private company LearnRank, an American company devoted to developing machine-learning algorithms that process data and evaluate professors around the world. According to the system, the rankings will obey a series of parameters. Three of them are especially relevant: the evaluation conducted by the Ministry in the last 5 years; the student's performance in national exams; and the student's feedback.*
>
> *In 2023, the Portuguese Government decides to apply such AI system to rank professors for the upcoming school year. When subject to the system, Maria ranks 'Poor' and receives a notification by the Ministry of Education arguing that she would be fired because of her evaluation.*
>
> *Maria tries to ask the government for explanations and the government refers to the law and the AI system. She then tries to get an explanation from LearnRank but the American company argues that property rights and industrial secrecy forbid it to disclose any kind of justification.*
>
> *Maria loses her job.*

This is the sort of example that illustrates the new challenges that may emerge in the future if we rely in AI without any precautions. In a classic administrative

relationship, the act of dismissing Maria's would be easily qualified as an administrative decision (the Portuguese *acto administrativo* or the French *acte administratif*). This would mean that Maria, according to most administrative systems in Europe (if not all), would be entitled to a justification or an explanation on the reasons why the Administration had chosen to let her go. The duty of the administration to give reasons for its decisions is, of course, a quintessential part of the rule of law in European liberal constitutional states.

The problem, however, is that in the example provided, the nature of the act in question becomes unclear, as public administration and private parties conflate to take the individual decision.

On the one hand, one could argue that the decision is an administrative decision. It ruled on a specific situation, changing Maria's legal position with it. It was practiced by a body of the public administration as it was its decision to fire Maria. On the other hand, however, such decision was a mere result (mandated by the Law) of the algorithm's output. This means that the Administration did not actually produce the reasoning leading up to its own decision, but it was rather a machine-learning system (a popular type of algorithm) that did it for it.

There are multiple administrative law questions one could pose regarding this case, but here are some to reflect on:

– Was this an administrative act if the ultimate decision was taken by the AI system?
– Was it a bounded or discretionary use of public power? Did the Administration have a real choice?
– Who is ultimately responsible for a faulty AI system? The Administration? The private contractor who built and operated the system? Both?

This set of questions lies at the core of administrative law. They represent the same interrogations we have been posing in the public law sphere every single day for centuries and do not represent any revolutionary rethinking. The answers, however, might radically differ from traditional approaches.

Take, for example, the nature of the act (question 1).

One must first consider whether such decision is a decision at all, meaning an act of will by the administration. Although this might seem a given, as every human decision is indeed an act of will, it is not so clear in AI decision-making. To qualify something as an administrative decision (*acte administratif*), one must first qualify such action by the administration as a legal act in itself, in some sort of ontological sense. Of course, with human reasoning all of this would be beyond the doubt. Traditional human decisions are indeed exercises of the will, the result of neural procedures that lead to certain outputs. Although we do not fully

understand such complex biological systems (e.g. how the human brain would reach certain ultimate conclusions), we can ask the decision-maker to try to elaborate a justification. In most cases, humans will be able to explain why they made a certain decision. In AI decision-making, however, this step is tremendously difficult. Most of these systems are 'black-boxes'[5] in which data is fed but whose process is often unknown or too complex to be explained. There is no exercise of human will, in the sense that no human decided on the final output, apart from the initial feeding of data. Of course, one might counterargue and say that there was human intervention along the way. A human decided to hire the AI company which developed the algorithm; another human decided on the code-language or the machine-learning method (supervised/unsupervised, for instance). Still, no human decided on the entire process of decision. No human was consulted during the process of enacting thar specific decision. It was rather the cooperation of public and private entities, of human and machine, that produced such an administrative decision.

Regarding the second question, one must understand whether an algorithmic decision should be considered as a discretionary or as a legally bound act. In the continental law tradition, this fundamental difference might shape the entire process and render such decision reviewable or unreviewable by the judiciary, guiding the level of scrutiny by the courts as well. One can see how both situations ought to be treated differently in the case of the use of algorithms. In the case of a discretionary act, the Administration has the freedom to choose between the many options. In these cases, the AI system will be used to help the administration in making a choice between such many possibilities and it will have the most fundamental role in shaping the decision. In the case of a legally bound act, however, the system is a merely automated procedure to expedite a decision which was bound to happen in any event. In such cases, the system merely expedited or automatized the decision, as a traditional machine would do in a factory setting.

Finally, regarding responsibility, the question is also very complex. According to Article 22 of the Portuguese Constitution, "Together with the officeholders of their entities and organs and their staff and agents, the state and other public entities are civilly liable for actions or omissions that are committed in or because of the exercise of their functions and result in a breach of rights, freedoms or guarantees or in a loss to others". If AI is used, there is no longer any officeholders of public bodies or civil servant who could be held liable for the administrative decision, and consequently the existing safeguards (in terms of disciplinary

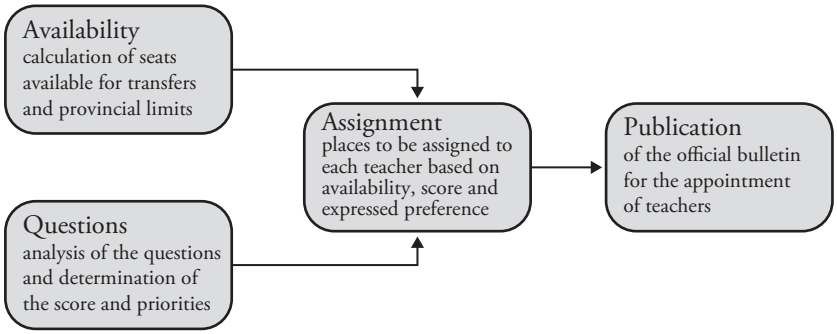
[5] F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (2015).

responsibility or civil liability) become inoperable. An algorithm or code cannot, in themselves, be held liable. What is left of those safeguards is the liability of the public entity (the state, the municipality or other public entity). Against this background, it can be asked what remains of the accountability categories (administrative, disciplinary, and tort liability) in a situation in which a decision-making power is transferred from a human decision-maker to an AI system.

This set of questions has been addressed in different situations across the many disciplines in which AI is being used by public administrations. We will now address some cases which illustrate citizen's vulnerability in face of public algorithmic administrations.

a) The Buona Scuola Case and Public Education

In 2015, the Italian Government decided to enact a law reforming the education system[6]. Among the many reforms, the government decided to automatize the process of allocation of professors in Italy. Its goal was apparently simple. To allocate new and old professors to existing vacancies based on a mix between professors' preferences and their objective merits via other past evaluations (including years of service). The system was developed by Hewlett-Packard (HP) Italia and Finmeccanica and it could be described graphically as such:



Source: La Sapienza/Tor Vergata

In 2016, the algorithm was put to the test and the results were disastrous. Many families were forced to move in opposite directions in Italy and all preferences

[6] Legge 13 luglio 2015 , n. 107, 'Riforma del sistema nazionale di istruzione e formazione e delega per il riordino delle disposizioni legislative vigenti', available at: https://docs.univr.it/documenti/Documento/allegati/allegati581858.pdf.

and merits seem to have been ignored by the system. This prompted a series of legal proceedings that discussed the matter of algorithmic decision-making by the public administration.

As Simoncini[7] puts it, once the case arrived in the Administrative Court of the Lazio region[8] (TAR Lazio), the administration's defense was twofold: i) first, it argued that the access to the software's source code did not fulfill the requirements to be considered an administrative act (not even a digital administrative act). This would mean that the individual professors were not entitled to see it or have someone analyze it. Second (ii), because the software had been created by a private company, the administration argued, the source code was protected by intellectual property rights and industrial secrecy. Both reasons would prevent the administration from giving any access to the source-code (and the explanations that go with it).

This constitutes one of the most common argumentations in algorithmic-based public procedures. As the administration does not possess the necessary know-how, it is forced to contract-out such technical expertise. By doing so, it can then skillfully argue that such action is not an administrative act, that there is not access to the source code because it is not an administrative document and that, in any event, the code is always protected by property rights.

Interestingly, the TAR Lazio Court was not convinced with this argumentation and dismantled the administration's reasoning. Although acknowledging the right to trade secrecy, the court clarified that the access to administrative documentation would trump the secretiveness of the algorithm in any case[9]. Moreover,

---

[7] Simoncini, 'Algorithmic Administration: The Constitutional Framework.', in Roberto Cavallo Perin and Diana-Urania Galetta (eds.), *Il Diritto Dell'amministrazione Pubblica Digitale* (2020) 30. pp. 9-10.

[8] Judgement of 22nd March 2017, 3rd Section B, Tribunale Amministrativo Regionale per il Lazio, judgement no. 2017/3769, available at: https://www.giustizia-amministrativa.it/.

[9] Judgement of 22nd March 2017, it reads: '- e, infatti, in materia di accesso agli atti della P.A., a norma dell'art. 24 della legge n. 241/1990, la natura di opera dell'ingegno dei documenti di cui si chiede l'ostensione non rappresenta una causa di esclusione dall'accesso;

- in particolare, la disciplina dettata a tutela del diritto di autore e della proprietà intellettuale è, come in precedenza brevemente rappresentato, funzionale a garantire gli interessi economici dell'autore ovvero del titolare dell'opera intellettuale, mentre la normativa sull'accesso agli atti è funzionale a garantire altri interessi e, in questi limiti, deve essere consentita la visione e anche l'estrazione di copia;

- né il diritto di autore né la proprietà intellettuale precludono la semplice riproduzione, ma precludono, invece, al massimo, soltanto la riproduzione che consenta uno sfruttamento economico e, non essendo l'accesso lesivo di tale diritto all'uso economico esclusivo dell'opera, l'ostensione deve essere consentita nelle forme richieste da parte dell'interessato, ossia della visione e dell'estrazione di copia, fermo restando che delle informazioni ottenute dovrà essere fatto un uso appropriato,

in a subsequent decision brought by different claimants, it seemed to forbid the administration from engaging in such delegation of public prerogatives (such as the allocation of professors) to 'robots' altogether[10]. According to the court, resorting to AI systems on discretionary decisions by the administration would be impossible as those automated systems lacked the 'technological procedures, [...] can never replace, truly supplanting it, the cognitive, acquisitive and judgmental activity that only an investigation entrusted to a natural-person-official is able to perform". Here it seems that the Court is not fully rejecting the use of AI systems in administrative action, but restricting its use to bounded administrative acts[11].

This case serves as an interesting first example of how AI procedures might infringe on fundamental rights and the role that administrative and constitutional principles might play on securing such rights.

b) The Loomis Case, COMPAS, and Public Justice

Another interesting example of vulnerability is that of criminal justice. Artificial intelligence systems are now increasingly used in criminal law and police settings, with different objectives. One of its most common uses is to assess criminal recidivism[12]. Criminal recidivism assesses the probabilities of a convicted criminal repeating any criminal conduct, hence aiding judges in sentencing and releasing timeframes. AI systems can of course provide interesting models to assess such recidivism probabilities as they manage to process the thousands of existing cases (through machine-learning processes) to produce a sort of recidivism risk rate which can help judges and prison authorities to decide on individual cases.

---

ossia esclusivamente un uso funzionale all'interesse fatto valere con l'istanza di accesso che, per espressa allegazione della parte ricorrente, è rappresentato dalla tutela dei diritti dei propri affiliati, in quanto ciò costituisce non solo la funzione per cui è consentito l'accesso stesso, ma nello stesso tempo anche il limite di utilizzo dei dati appresi, con conseguente responsabilità diretta dell'avente diritto all'accesso nei confronti del titolare del software.

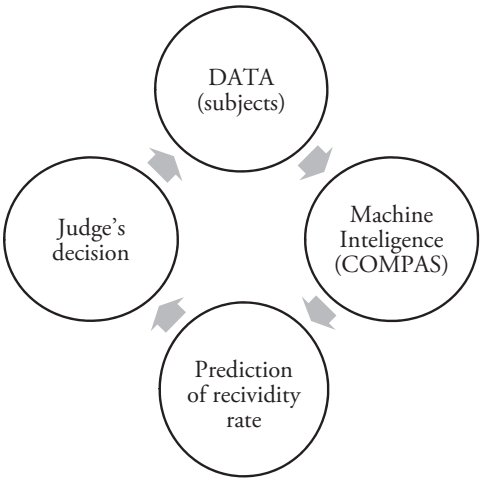
[10] Judgment of Setember 2018, Tribunale Amministrativo Regionale per il Lazio, 3rd Section B, joint cases no. 9224-9230 – para. 6 - Invero Il Collegio è del parere che le procedure informatiche, finanche ove pervengano al loro maggior grado di precisione e addirittura alla perfezione, non possano mai soppiantare, sostituendola davvero appieno, l'attività cognitiva, acquisitiva e di giudizio che solo un'istruttoria affidata ad un funzionario persona fisica è in grado di svolgere.

[11] Simoncini, *supra* note 7. p. 10.

[12] Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 Notre Dame L. Rev. (2016). Available at: https://scholarship.law.nd.edu/ndlr/vol91/iss2/2.

A famous example of one such system is the so-called Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system. The COMPAS system[13] consists of an advanced AI system, developed by Northpointe, which aims at generating risk scales to help judges deciding on both pre-trial sentencing (defining, for example, the risk of a certain individual committing felonies pending trial) and post-trial assessments (e.g. the risk that a certain prisoner will commit further violent acts). This means that judges are given a risk scale which is supposed to guide part of their judgment in individual cases.

Just like any AI system, including the ones described above, the COMPAS system makes use of enormous amounts of data to generate such predictions. Graphically it could be described as below:



According to its explainer guide[14], the system gives a certain weight to a set of parameters which include:

- History of Noncompliance Scale
- Vocational Education Scale
- Current age
- Age-at-first arrest
- History of Violence Scale

[13] Practitioner's Guide to COMPAS Core, p. 29. Available at: https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf.

[14] Practitioner's Guide to COMPAS Core, available at https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf.

The number (weight) given to each parameter informs a methodology which can be described as such:

Violent Recidivism Risk Score = (age∗–w)+(age-at-first-arrest∗–w)+(history of violence∗ w) + (vocation education ∗ w) + (history of noncompliance ∗ w)

This equation would then produce a risk score which should help judges in deciding individual cases[15]. This system is now currently in use in different states of the United States of America, but many other examples can be found across Europe.

In 2016, the Supreme Court of Wisconsin was asked to rule on the matter of the legality of the use of COMPAS in that state's judicial system[16]. The facts of the case were as such. In early 2013, Wisconsin charged defendant Eric Loomis with five criminal counts related to a drive-by shooting. Loomis then pleaded guilty to some of the minor accounts and was sentenced to six years of imprisonment and five years of extended supervision[17]. In the sentencing procedure, the Wisconsin court relied on COMPAS to assess Loomis recidivism and the AI system helped framing the number of years that Loomis would serve. Loomis then argued that the use of the system constituted a breach of his due process rights because there was no possibility of accessing the algorithmic behind the system as this was labelled as 'trade-secret'. Moreover, the defendant also argued that the system was discriminatory, claiming that its predictions were biased against male defendants, again breaching his constitutional rights to due process.

Interestingly, the Wisconsin Supreme Court did not rule in favor of Loomis but rather defended the legality of the COMPAS system. The core argument of the court was that judges retained full discretion to consider COMPAS risk assessments and had hence full freedom to disregard them. This meant, *a contrario*, that COMPAS systems should never be used as the only factor in the decision to convict a person and caution should be taken as to its influence in decision-making procedures. Apart from those cases, the system would be otherwise lawful.

In any event, the Wisconsin Supreme Court went further and specified some warnings that should accompany COMPAS results. According to the Court, the reports that accompanied the COMPAS risk assessment should have at least i)

[15] See the New York State New York State COMPAS-Probation Risk and Need Assessment Study: Examining the Recidivism Scale's Effectiveness and Predictive Accuracy https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=269445.

[16] Loomis case - State of Wisconsin, Plaintiff-Respondent, v. Eric L. Loomis, Defendant-Appellant. https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690.

[17] For a summary of the case see https://harvardlawreview.org/2017/03/state-v-loomis/.

written warnings to judges alerting to the proprietary nature of COMPAS system, meaning the trade-secret attached to the system, ii) a warning that COMPAS was based on group data, and not the individual specific situation; iii) information clarifying that COMPAS' data was nationally gathered, and did not reflect the State's specific population; iv) information on the studies which have raised doubts about whether the COMPAS system disproportionately classified minority offenders with a higher risk; and v) a final disclaimer that the system only worked as an advisory tool in *post-sentencing* and not as a tool to actual determine whether a certain person ought or ought not to be convicted of a crime.

This case raised many concerns, especially given the decision to uphold Northpointe's right to trade-secrecy. One of the most alarming studies was conducted in 2016 by the NGO ProPublica, in an extensive analysis of the AI system which COMPAS used[18]. In such study, Propublica concluded the following[19]:

i) Black defendants were often predicted to be at a higher risk of recidivism than they actually were. The analysis concluded that black defendants who did not recidivate over a two-year period were nearly twice as likely to be misclassified as higher risk compared to their white counterparts (45 percent vs. 23 percent).

ii) White defendants were often predicted to be less risky than they were. The analysis concluded that white defendants who re-offended within the next two years were mistakenly labeled low risk almost twice as often as black re-offenders (48 percent vs. 28 percent).

iii) The analysis also showed that even when controlling for prior crimes, future recidivism, age, and gender, black defendants were 45 percent more likely to be assigned higher risk scores than white defendants.

iv) Black defendants were also twice as likely as white defendants to be misclassified as being a higher risk of violent recidivism. And white violent recidivists were 63 percent more likely to have been misclassified as a low risk of violent recidivism, compared with black violent recidivists.

v) The violent recidivism analysis also showed that even when controlling for prior crimes, future recidivism, age, and gender, black defendants were 77 percent more likely to be assigned higher risk scores than white defendants.

---

[18] Julia Angwin *et al*, "Machine Bias", ProPublica, 23 May 2016, available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[19] Jeff Larson *et al*, "How We Analyzed the COMPAS Recidivism Algorithm", ProPublica, 23 May 2016, available at: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

These studies show the importance of the administrative principle of due process and transparency. While in the Loomis case the Court held Northpointe's claim that its algorithm constituted trade-secrecy, it is important to understand the impact that such decision might have on public decisions and the creation of vulnerability for citizens. Not being able to review the exact way a certain decision was made, even if not the final administrative act (the sentencing act), might still constitute a true impossibility of judicial redress to some citizens. This was in fact what ultimately led to TAR Lazio rejecting the use of such AI systems in discretionary acts. One would not expect such opacity in normal administrative action, and courts would probably deem it as a lack of administrative justification. Why accept it within the AI context? Traditional administrative principles of due process and non-discrimination must still be respected, even in face of new technologies, and must be properly balanced against other fundamental rights such as property.

c) Weapons of Math Destruction, Predpol and AI Police

Both cases above represent potential candidates to what Cathy O'Neil calls Weapons of Math Destruction[20] (WMD). This concept describes those algorithms which shape how public administration delivers decisions in such a way that they end-up creating loops of systematic breaches of fundamental rights. As seen in both the *Buona Scuola* and the COMPAS cases, both algorithms delivered effective breaches of fundamental rights, albeit being judged in different ways by the respective courts. Cathy presents us the bigger picture, by showing the impact of such tools in terms of equality and class discrimination.

A good example of how WMD might have a strong impact on fundamental rights, is how AI systems systematically disregard less represented minorities in technologies such as facial recognition. According to a recent study[21], facial recognition systems consistently underperform in identifying dark-skinned individuals, including those systems from tech behemoths like IBM or Microsoft. Although most of the cases of discrimination have been dealt within the private sector, including Amazon's famous gender-biased hiring system[22], the problem

[20] C. O'Neil, *Weapons of math destruction: how big data increases inequality and threatens democracy,* New York Crown Publishers, (2016).
[21] http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.
[22] The Guardian, "Amazon ditched AI recruiting tool that favored men for technical jobs", 11/11/2018, https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-

extends to public administrations. As citizens data is more and more transferred to public databases – for example for security purposes – problems might arise if the technology is not fed the right data.

For example, in the United States, it has become clear that police's uses of AI facial recognition systems tend to target black people over white people, based on the collected data[23]. Moreover, it often fails to correctly identify black people, leading to situations in which innocent people are incorrectly targeted and persecuted based solely on a faulty AI system. Amnesty International has reported dozens of cases of such failures, including the case of Derrick "Dwreck" Ingram[24] in which NYPD used his personal Instagram account to use facial recognition systems in the city to track him down. This kind of activity has led some cities in the U.S to simply ban this kind of technology when used in police investigations.

Sometimes, however, it is not the system itself that breaches fundamental rights, but the fact that it creates dangerous confirmation loops. A good example of this is how police departments are using geo-location AI systems to allocate police officers across big cities.

The idea seems rather innocuous. To create a system that geolocates the areas with more criminality rates and that then predicts where is more likely for crime to occur next, hence allowing for better personal allocation. One of the most famous examples of such algorithms is a system called Predpol[25] (New York uses a similar system called CompStat[26]), which aims at predicting where big pockets of crime can exist in cities like Atlanta. The problem with this type of AI system is that it might lead to the so-called feedback loops[27]. When these systems are made to target every type of crime (not only serious crime) they will tend to point more towards the poor (and often minority-resident) parts of the cities. This is simply because it is in those areas where small crime tends to occur more often. Once more police is allocated to those places, the tendency is naturally

---

engine . See also the fictional case Wicks, Budd, Moorthi, Botha and Mead 'Automated Hiring at Amazon', Darden Case No. UVA-E-0470, SSRN, 2021.

[23] Buolamwini and Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', 81 *Proceedings of Machine Learning Research* (2018) 15.

[24]Amnesty International: "Ban dangerous facial recognition technology that amplifies racist policing", 26/01/2021, https://www.amnesty.org/en/latest/news/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/.

[25] https://www.predpol.com/.

[26] https://www1.nyc.gov/site/nypd/stats/crime-statistics/crime-statistics-landing.page.

[27] De Abreu Duarte, 'AI, Hume and a guillotine: The dangers of machine-learning loops', EUI Ideas, 2021, available at: https://euideas.eui.eu/2021/04/21/ai-hume-and-a-guillotine-the-dangers-of-machine-learning-loops/.

for small burglary and small crime to be more caught, hence feeding the system with more confirmation that more police is needed. This feedback loop, confirming the initial suspicions of the AI systems, can grow to a point in which becomes empirically clear that poor people are criminals simply because they are poor. This does not mean that other areas of the city do not have an equal amount of crime; there are simply less policemen around to catch it and so the data fed into the system ignores those. From that point onwards, it is not difficult to politically exploit such data. The normative point then becomes that poor people are criminals.

The same problems we saw in the education and justice sectors are in play here. Often times, these systems are not developed by police departments but rather by private contractors. This means that once a certain detainee claims due process or algorithmic transparency, chances are that the same argumentation of the *Buona Scuola* or the *Loomis* arises in Court. However, it does not necessarily have to be like this. Although opacity seems to be the rule in the development of such systems, some examples prove that it is possible for algorithmic administrations to do better in terms of transparency[28]. The experience with the Durham Constabulary's Harm Assessment Risk Tool (HART)[29] – a similar tool to COMPAS deployed in the United Kingdom – is a good example on how public AI systems might be transparent to the general public[30].

d) Centrelink and Social security

The Centrelink Debt Program was established in July 2016 by Centrelink, within the Department of Human Services (DHS) of Australia. It commenced using a new online compliance intervention (OCI) system, which was automated, for raising and recovering debts. The use of this system has been often referred to as 'Robodebt', primarily because the system led to debt recovery letters being automatically generated by a computer program.

---

[28] Oswald, 'Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power', 376 *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (2018). p.4.

[29] See Durham Constabulary's AI decision aid for custody officers, A case study on the use of AI in government, Centre for Public Impact, 2018, available at: https://www.centreforpublicimpact.org/assets/documents/ai-case-study-criminal-justice.pdf.

[30] Oswald et al., 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality', 27 *Information & Communications Technology Law* (2018) 223.

In its 2017 inquiry into the OCI system, the Australian Commonwealth Ombudsman described the operation of the automated system as follows: "The OCI matches the earnings recorded on a customer's Centrelink record with historical employer-reported income data from the Australian Taxation Office (ATO). Parts of the debt raising process previously done manually by compliance officers within DHS are now done using this automated process. Customers are asked to confirm or update their income using the online system. If the customer does not engage with DHS either online or in person, or if there are gaps in the information provided by the customer, the system will fill the gaps with a fortnightly income figure derived from the ATO income data for the relevant employment period ('averaged' data)".[31] The deployment of the Centrelink Debt Program, which used this OCI system, resulted in a huge increase in the scale of DHS's debt-raising and recovery process.

When the Centrelink Debt Program commenced operation, it became apparent that the 'averaged' data process was resulting in the generation of inaccurate debt notices, which has a particular impact on a number of recipients who were already marginalized.[32] Both the Commonwealth Ombudsman's and the Senate Standing Committee on Community Affairs' respective inquiries identified that many of the problematic aspects of the Centrelink Debt Program related to how the system was rolled out. This included, for example, the lack of information, and difficulty accessing information about how to challenge or seek review of a debt nominated in a debt recovery letter.[33]

In November 2019, the Australian federal government settled a landmark challenge to its 'Robodebt' program, conceding that income-averaging alone was not a proper basis to claim that somebody owed it a debt. In orders made by

---

[31] Australian Commonwealth Ombudsman Centrelink's automated debt raising and recovery system: A report about the Department of Human Services' Online Compliance Intervention System for Debt Raising an Recovery (April 2017), 1 at http://www.ombudsman.gov.au/__data/assets/pdf_file/0022/43528/Report-Centrelinksautomated-debt-raising-and-recovery-system-April-2017.pdf.

[32] Errors and discrepancies arose when an assumption was made about income, and, consequently, incorrect information being included in the OCI's calculation: see Senate Standing Committee on Community Affairs *Design, scope, cost-benefit analysis, contracts awarded and implementation associated with the Better Management of the Social Welfare System initiative* (21 June 2017), [2.85]-[2.101], available at https://www.aph.gov.au/parliamentary_business/committees/senate/community_affairs/socialwelfaresystem/Report/c02.

[33] See Senate Standing Committee on Community Affairs 'Chapter 3 – Communicating with Centrelink' *Design, scope, cost-benefit analysis, contracts awarded and implementation associated with the Better Management of the Social Welfare System initiative* (21 June 2017), available at https://www.aph.gov.au/parliamentary_business/committees/senate/community_affairs/socialwelfaresystem/Report/c02.

consent – agreed by the commonwealth – the Federal Court of Australia declared the plaintiff's debt was "not validly made", that an order to garnish her tax return was also invalid and there was no basis to add a 10% penalty to the debt.[34] The Court said the conclusion the plaintiff had received social security benefits she was not entitled to was "not open on the material before the decision-maker" because there was "no probative material" that the average reflected her actual income. "In the circumstances, there was no material before the decision-maker capable of supporting the conclusion that a debt had arisen (...) The conclusion that a debt had arisen was therefore irrational, in the requisite legal sense."

The United Nations Special Rapporteur on Extreme Poverty and Human Rights, Philip Alston, in his warning about the risk of a "digital welfare dystopia," singled out Robodebt as one of the leading examples of how much human and reputational damage can be caused by bad design.[35]

This is one example of the possible dangers of AI and automation in this area. The digital administration is commonly presented as a politically neutral enterprise designed to ensure that citizens benefit from new technologies, experience more efficient government, and enjoy higher levels of well-being. However, it has been often used to promote reductions in the overall welfare budget, a narrowing of the beneficiary pool, and the increase of control and surveillance of the beneficiaries. It is necessary to ensure that these new digital welfare systems abide by human rights standards and follow due administrative procedure principles.

Section II – Solving the Problem: principles for the Algorithmic Administration

Now that we have briefly understood some of the dangers posed by the algorithmic administration, it is time to give the first steps towards its governance. We suggest that the problems which we face today, albeit new, can be tackled through existing administrative reasonings. These require, however, an updated view on the relationship between administration and citizens.

If one analyzes the problems described through the lens of classic administrative law, it becomes clear that the problems faced are similar to those discussed in the

[34] Order of the Federal Court of Australia, District Registry: Victoria, Division: General, No: VID611/2019 Deanna Amato v. The Commonwealth of Australia, 27 November 2019, https://www.comcourts.gov.au/file/Federal/P/VID611/2019/3859485/event/30114114/document/1513665.

[35] "World stumbling zombie-like into a digital welfare dystopia, warns UN human rights expert", 17 October 2019, https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25156.

early times of Administrative Law. One the one hand, we have an opaque administration, whose actions are neither justified nor explained to the citizens. This is the case of when a professor is poorly allocated, or a criminal wrongly identified, or someone receives the notification of a social security debt, and no justification is given apart from 'the AI said so'. It does not differ so much, after all, from the old 'the King said so'. Here we see how core principles of administrative action such as due process or accountability might need to be urgently updated to encompass the technological dimensions of this new *modus operandis*. On the other hand, we face problems of responsibility which resemble those of an 'immune' (unaccountable) administration, typical of nineteenth century regimes more than twenty-first century liberal constitutionalism.

The difference however – and this is an important point – is the role of privates in designing the system. While in all-powerful public administrations of the past, based on royal or dictatorship prerogatives, the violation of fundamental rights came directly from the administration, now this is no longer the case. Today, the administration shields itself on the privates' rights of intellectual property or trade secrecy, hence deflecting the need for administrative transparency. Simultaneously, the private argues that the ultimate decision was that of the administration and that the software merely was designed as a mere aiding tool. Through property rights and this *pingpong* responsibility, the citizen is the one left unprotected.

In any event, the fundamental questions remain the same. AI systems must be subjected to the rule of law[36] just like any other instrument at the service of the administration. In a systematic way, one could argue that the principle of the rule of law empowers citizens by allowing them to ask three fundamental questions to the administration: the *i) what,* the *ii) why and* the *iii) who* of administrative action. These remain at the core of algorithmic transparency.

We suggest to connect those questions to four interdependent principles for the algorithmic public governance of the future[37]: transparency; due process; responsibility and non-discrimination.

Firstly, any public administration, algorithmic or not, must explain the *what* of its actions. This means that the administration must allow citizens to see what it does, by granting them access to their documents and files. The algorithmic administration of the future must be *transparent*. If systems are opaque, in the

[36] Hildebrandt, 'Algorithmic Regulation and the Rule of Law', 376 *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (2018).
[37] Adopting a different set of principles see Coglianese and Lehr, 'Regulating by Robot: Administrative Decision Making in the Machine-Learning Era', 105 *The Georgetown Law Journal* 78. p. 1176.

sense that they are simply not scrutable (like so many of the AI systems), then the power of the administration cannot be checked by the people and we lose a fundamental achievement of liberal constitutionalism.

This, however, will not be sufficient. Transparency in AI systems is a complicated business and requires some thinking 'inside' of the box[38]. This means looking at the procedures within the AI system (what Koivisto calls logic of discovery[39]) but also make it understandable to citizens (Koivisto's logic of justification[40]). Most of machine-learning systems, even if scrutinized, would provide very little on how they work. By drawing their own parameters from big sets of data, mere administrative transparency (as, for example, accessing administrative documentation or looking at the algorithm) will most likely not inform any citizen. The complexity of the systems, attached to the obstacles posed by private companies, renders any exercise of mere transparency a tortuous path for individuals. This is why transparency, by itself, is not enough.

Hence, the algorithmic administration must also explain the *why* of its actions. This means that citizens must be informed of the reasons that led the administration to adopt a certain decision. The algorithmic administration of the future must be *justified and explainable*. This is a fundamental aspect of the rule of law, portraying the change from the absolute administrations of the past, to those modern which derive power directly from the people and that must respond to them. Justification and explainability are normally one and the same in classic administrative action. When a citizen asks the administration to justify a given administrative decision, it ought to receive an explanation, in natural language (hopefully in that of the citizen), of why such legal act was performed. The justification and explainability are part of the same reasoning as provided by the administration. In algorithmic procedures, however, the distinction is of the utmost importance. Although one could simply justify a certain act with reference to the AI system (e.g. the AI decided X under this algorithm and that is why you were sent to school Y), this does not suffice to meet the explainability threshold. Explainability requires the administration not only to give access to the complex algorithmic data that served as reasoning for the administrative act, but also to render it understandable to the common citizen. Explaining AI decision-making will be a fundamental part of the duty to provide reasons in algorithmic administrations.

---

[38] Koivisto, 'Thinking Inside the Box: The Promise and Boundaries of Transparency in Automated Decision-Making', (2020) 29.
[39] *Ibid.* pp. 17-19.
[40] *Ibid.*

The question of when and what kind of explanation might be required of AI systems is discussed under the general term of the "right to explanation". This is enshrined in the European Union's General Data Protection Regulation (GDPR) that provides a right to information about the existence, logic, and envisaged consequences of automated decision-making systems in Articles 13 through 15, as well as a right not to be subject to automated decision-making processes in Article 22.[41] The degree to which this constitutes a "right to explanation" is the subject of significant debate, this unarguably prompts the question of whether explanations are necessary or sufficient to convey meaningful information about the operation of AI and other automated decision-making systems.[42]

The issue of explainable AI has been noted by a number of public bodies in the United States[43], the United Kingdom[44], and France.[45] While there is significant support for explanations as a tool for holding AIs accountable, there are also concerns about the costs of generating explanations, the engineering challenges surrounding explanation from AI systems would stifle innovation, or that explanations might force trade secrets to be revealed. The delicate balancing between all of these conflicting values should not, however, hinder citizens from obtaining meaningful explanations. Solutions exist, including explainer algorithms[46], that might help decoding such complexity.

Thirdly, the modern public administration must be *held accountable* to citizens and its actions must be judicially reviewable. This means that citizens must be

---

[41] Council Regulation 2016/679, arts. 13-15, 22, 2016 O.J. (L119) 1.

[42] *Finale Doshi-Velez et al* Accountability of AI Under the Law: The Role of Explanation https://arxiv.org/abs/1711.01134 .

[43] DARPA, Broad Agency Announcement, Explainable Artificial Intelligence (XAI), DARPABAA-16-53, at 6 (August 10, 2016), available at https://www.darpa.mil/attachments/DARPABAA-16-53.pdf.

[44] See House of Lords, Select Committee on Artificial Intelligence, Report of Session 2017-19, AI in the UK: Ready, Willing, and Able? (April 16, 2018), available at https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf; House of Commons, Science and Technology Committee, Fourth Report of Session 2017-19, Algorithms in Decision-Making (May 15, 2018), available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf.

[45] See France Intelligence Artificielle, Rapport de Synthèse (Jan. 2017), available at https://www.economie.gouv.fr/files/files/PDF/2017/Rapport_synthese_France_IA_.pdf (French only); Cédric Villani, For a Meaningful Artificial Intelligence: Towards A French and European Strategy (March 28, 2018), available at https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.

[46] Ribeiro, Singh and Guestrin, "Why Should I Trust You?': Explaining the Predictions of Any Classifier', *ArXiv:1602.04938 [Cs, Stat]* (2016), available at http://arxiv.org/abs/1602.04938 (last visited 10 May 2021].

granted the possibility of appealing to the judiciary to check the power of the administration, holding it responsible for breaches of the law. They must understand *who is responsible* for a given administrative decision. This corollary is particularly important when most of the algorithmic action is designed by private actors, many times under complex AI systems whose parameters' control is not completely in the hands of any creator. It is not acceptable for the administration to shield in private property to avoid its own responsibilities. If the ultimate decision is made by the administration, even if mandate by an automated decision by an AI system, it should the administration to bear the responsibility for these actions.

When the AI system is developed by private companies, especially when such companies refuse to provide access to the source-code/machine learning methods, then the citizen must be entitled to seek redress with the administration regardless of the obstacles of the private. This means to create a preferential right by which the citizen is always entitled to pursue legal action against public administration first and to obtain compensation for the damages from it. Whether or not there is then a subsequent compensation by the private should be up for the administration to solve, not the citizen. This governance of responsibility is proportional to the dimension of information that all players involved possess. The citizen is the weakest party in the entire procedure and should be, especially as a consumer, entitled to the utmost protection.

Finally, the algorithmic administration must be bound to the principle of equality and non-discrimination. This means two things. First, that the algorithmic administrative state sometimes gives beneficiaries the option to choose between a digital solution or to continue using more traditional techniques. In reality, however, policies such as 'digital by default' or 'digital by choice' are usually transformed into 'digital only' in practice. This in turn exacerbates or creates major disparities among different groups. A lack of digital literacy leads to an inability to use basic digital tools at all, let alone effectively and efficiently. Limited access, or no access to the internet, as well as the costs of access to the internet, poses huge problems for a great many people. Second, that when an AI system is used to help in public decision-making, the administration is bound to ensure that such system complies with the principle of non-discrimination. Systems such as the ones described above, which make use of facial recognition technology, cannot be tolerated for public use if biased in favor of certain majority groups. It is our view that, if such case arises due to a faulty AI system (even if developed by a private company), it is for the administration to respond in first place. This is the only way to safeguard citizens' fundamental rights in face of the difficult paths of discovery and transparency we described before.

All of these principles are procedural in nature and aim at ensuring that citizens can have a voice against the power of the algorithmic administration. Only by empowering citizens to question and scrutinize public decisions, might we be able to ensure full accountability of algorithmic public decisions.

Conclusions

This paper has briefly introduced the problem of AI administrative decision-making procedures and posed some of the questions we will be facing in the future. We believe it is not a matter of if, but rather of when, we will be faced with fully automated procedures in many areas of administrative action. This will bring about challenges to fundamental rights and test our systems' capacity to react to novel technologies.

AI systems and automated decision-making are not good or bad *per se*. They do not represent a silver bullet that will make all administrative decisions impartial and correct, in the same way that and they are not an evil black box that will rule us all. They aim at making things easier, faster, and more cost-efficient. They simply represent a different way of deciding and come with both advantages and disadvantages. The risk, however, is that by not preparing ourselves for its arrival, we jeopardize its correct use and will not profit from its efficiency gains. If public administrations do not avoid the possible pitfalls of the incorrect adoption of such systems, chances are that it will be for the judiciary to establish the necessary corrections when reviewing its actions. This might then result in an absolute rejection of such systems (*Buona Scuola*) or an excessive protection of IP and trade-secrecy rights (*Loomis*). In any event, it would be better for the administration to foster open and transparent AI systems, under the scrutiny of the public, with full respect for the due process tradition.

We conclude that the principles of transparency, due process, accountability, and non-discrimination might serve as the first cornerstones of the algorithmic administrative law of the future. Given the complexity of the situations in which AI plays a role, as well as the highly technical dimension of its systems, we defended that administrative principles should be of a procedural nature, empowering citizens to know more and obtain justifications from the administration. Only by giving citizens a voice can we expect both public administrations and privates to open the black boxes of AI systems and let some transparency shine inside.