7

A Tale of Two Automated States

Why a One-Size-Fits-All Approach to Administrative Law Reform to Accommodate AI Will Fail

José-Miguel Bello y Villarino

7.1 INTRODUCTION: TWO TALES OF THE AUTOMATED STATE

In his 1967 book, which partially shares its title with this edited collection (*The Automated State: Computer Systems as a New Force in Society*),[1] Robert McBride anticipated that public authorities would be able to do 'more' thanks to the possibility of storing more detailed data combined with the increasing capacity of machines to process that data. He conjectured that this would create new legal problems. Fast forward half a century and the Automated State may (really) be on the brink of happening. AI can essentially change the state and the way it operates – note the 'essentially'.

Public authorities, employing (or assisted by) machines to a large scale, could do more. What this 'more' is, is a matter of discussion,[2] but, broadly speaking, it can mean two ideas: (i) doing things that humans could do, but more efficiently or to a larger scale; or (ii) doing things that could not be done before, at all or at a reasonable cost.[3] Therefore, the rules that regulate the action of public authorities need to be adapted. This chapter deals with the normative question of the type of regulatory reform that we should aim for.

It can be anticipated that changes within the immediate horizon – three to five years – will be marginal and starting at the points of least resistance, that is, in tasks

[1] Robert McBride, *The Automated State: Computer Systems as a New Force in Society* (Chilton Book Company, 1967).

[2] WG de Sousa et al, 'How and Where Is Artificial Intelligence in the Public Sector Going? A Literature Review and Research Agenda' (2019) 36 *Government Information Quarterly* 101392; BW Wirtz, JC Weyerer, and C Geyer, 'Artificial Intelligence and the Public Sector – Applications and Challenges' (2019) 42 *International Journal of Public Administration* 596–615.

[3] K Gulson and J-M Bello y Villarino, 'AI in Education' in Regine Paul, Emma Carmel, and Jennifer Cobbe (eds), *Handbook on Public Policy and Artificial Intelligence* (Edward Elgar, forthcoming 2023).

https://doi.org/10.1017/9781009334297.011 Published online by Cambridge University Press

currently done by humans that could be easily automated. In these cases, the preferred regulatory option is likely to be the creation of some *lex specialis* for the situations when public authorities are using AI systems. This approach to automating the state and the necessary changes to the administrative law are explored in the following section (Section 7.2).

The much bigger challenge for the regulation of the Automated State will come from structural changes in the way we design policy and decide on policy options. This is best illustrated with one example already in the making: digital twins, data-driven copies of existing real-life environments or organisms. Although the attention has primarily focused on digital twins of living organisms,[4] promising work is being undertaken in other types of real-life twins, such as factories or cities. One leading example is the work in Barcelona (Spain) to create a digital twin that will help make decisions on urban policy, such as traffic management or planning.[5]

According to some reports, when one of the key planning initiatives of the local government – the *superilles*, which involved the creation of limited-traffic city-block islands – was run through the system to see the effects with and without its implementation, it showed that there was close to no improvement on air pollution levels, one of the drivers for the creation and implementation of the initiative.[6] In other words, the intervention failed to achieve one of its main goals. Does this matter for administrative law?

Section 7.3 considers these policy-oriented types of AI systems. The systems used to design policy and make decisions among policy options open the door to an intrinsically different automated state which may require completely new tools and approaches to regulate it. Although the word 'automated' could be misleading – it is better described by the periphrastic 'AI-driven decision support system for policy design and creation' – the outputs of these systems are within the scope of administrative law. They are part of processes that eventually generate administrative acts or decisions and, as such, can be the object of challenges on legal grounds in many jurisdictions.

A key part of that discussion is the problem of translating into law a procedure for legal administrative accountability for 'objectives' (a particular type of input for those AI systems) and 'insights' (outputs). AI systems are often developed to optimise a number of objectives set by humans or to autonomously find insights and interesting relations among the data fed into it. When these types of AI systems are used on data held by public authorities for policy-making purposes, they generate immediate

[4] S Scoles, 'A Digital Twin of Your Body Could Become a Critical Part of Your Health Care' (10 February 2016) *Slate*; J Corral-Acero et al, 'The "Digital Twin" to Enable the Vision of Precision Cardiology' (2020) 41 *European Heart Journal* 4556–64.

[5] J Argota Sánchez-Vaquerizo, 'Getting Real: The Challenge of Building and Validating a Large-Scale Digital Twin of Barcelona's Traffic with Empirical Data' (2022) 11 *ISPRS International Journal of Geo-Information* 24.

[6] A Hernández Morales, 'Barcelona Bets on "Digital Twin" as Future of City Planning' (18 May 2022) *Politico*.

challenges to administrative law: how do we regulate policy-making that is meant not to be about discretionary choices, but about data-driven optimisation?

Concepts such as 'arbitrariness' or 'discretion' mean very different things in regard to public authorities' decisions which are an application of the law to individuals or groups, covered in Section 7.2,[7] and for decisions about how to best use public resources at a policy level, explored in Section 7.3.[8] Distinguishing between legitimate political (or policy) choices and unreasonable decisions will be challenging if at a given stage of the decision-making process there is a system that is considering one option preferable to another according to the parameters built into that system.

This type of problem may still be incipient. The technology may still be very far from reliable, but if we reach a point when some policies can be shown to be Pareto superior to others (i.e., not one of the indicators considered in the policy is worse-off, but at least one is improved), is the choice of the Pareto inferior option still legitimate or fair? Will it be legal? How much deference should then be given to the choices of decision-makers?

To solve some of these questions in Section 7.4, I suggest some preparatory work for this scenario. I develop some heuristics – or rules of thumb – to distinguish between both tales of the Automated State. On that basis, I explore whether democratic and liberal societies can create a new type of administrative law that can accommodate divergence of views and still ensure that the margin of discretion of policy choices is adjusted to this new reality.

7.2 THE ADMINISTRATIVE LAW OF AI SYSTEMS THAT REPLACE BUREAUCRATS

The use of AI for automating work currently done by humans – or creating systems that facilitate the performance of those tasks by humans – can be directly linked to previous investments by governments in information systems. These were generally associated with attempts to update the ways public organisations operated to enhance efficiency and policy effectiveness.[9] Those AI systems, if used for fully automated administrative tasks, could be 'isolated from the organisational setting they originated from'[10] and, therefore even legally considered as 'individual artificial bureaucrats'.[11]

[7] See, for example, the discussion about discretion in different levels of bureaucracy in JB Bullock, 'Artificial Intelligence, Discretion, and Bureaucracy' (2019) 49 *The American Review of Public Administration* 751–61.

[8] See also the discussion in Chapter 10 in this book.

[9] See A Cordella and N Tempini, 'E-Government and Organizational Change: Reappraising the Role of ICT and Bureaucracy in Public Service Delivery' (2015) 32 *Government Information Quarterly* 279–86 at 279, and the references therein.

[10] Ibid, 281.

[11] JB Bullock and K Kim, 'Creation of Artificial Bureaucrats' (Lisbon, Portugal (Online), 2020), 8.

In this context, the main consideration is that the system should be able to do its job properly. This view, therefore, naturally places the accent on testing the AI systems beforehand, particularly for impartiality and standardisation. This is something we are relatively familiar with and not conceptually dissimilar to the way Chinese imperial mandarins were subject to excruciating exams and tests before they could work for the emperor, or to the way the Spanish and French systems (and the countries in their respective areas of influence) still see the formalised gruelling testing of knowledge as a requisite to access a 'proper' bureaucrat position.

Therefore, administrative rules for the use of these AI systems are likely to focus on the systems themselves. As mentioned, the regulatory approach will then most likely emphasise ensuring that they are fit for purpose before starting operation, which is a type of legal reform already observed in several jurisdictions.

Commonly cited examples are the mechanisms already in place in Canada,[12] which focus on the risks of AI systems employed by public authorities; the proposed general approach in the European Union,[13] which expands to high-risk systems in the public and private sector; or the light-touch intervention model, which creates some pre-checks for the use of certain AI systems by the public authorities, such as the recently introduced rules in the state of New South Wales in Australia[14] – although with no concrete consequences, in this case, if the pre-check is not done properly.

Generally speaking, these approaches place the stress on the process (or its automated part) and not on the outputs. It is the system itself that must meet certain standards, defined on the basis of actual standards or specifications (in the EU case as described in article 9 of the proposal) or an impact assessment of some kind (Canada model) or the considerations of 'experts' (New South Wales, Australia model). At a higher level, this makes sense if what we are concerned about is the level of risk that could be generated by the system. The question here is 'how bad can it go?', and the law mandates to undertake that check beforehand.

In my opinion, this deviates from the views of administrative law that see the action of the public authorities as a materialisation of values such as equality and fairness.[15] Instead, this Weberian machine bureaucracy would stress impartiality and standardisation,[16] values more intrinsically attached to procedural elements.[17]

[12] Treasury Board of Canada, *Directive on Automated Decision-Making* (2019).

[13] European Commission, *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts* (Proposal, 21 April 2021), see also Chapter 1 in this book.

[14] Digital.NSW, NSW Government, *NSW AI Assurance Framework* (Report, 2022).

[15] S Verba, 'Fairness, Equality, and Democracy: Three Big Words' (2006) 73 *Social Research: An International Quarterly* 499–540.

[16] TM Vogl et al, 'Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities' (2020) 80 *Public Administration Review* 946–61 at 946.

[17] See also discussion in Chapter 12 in this book.

In the classic model of Peters, in which the public administration is a manifestation of a combination of societal, political, and administrative cultures,[18] the direct connection here is to the administrative culture, and only collaterally to societal or political elements. That type of Automated State does not need to be fair, it needs to be accurate. The fairness is meant to be embedded in the policy it implements and the legitimacy of outputs depends on whether the process correctly implements the policy.

However, as this approach incorporates elements of risk-based regulatory techniques, outputs are indeed considered in the process of conformity checks. Normally, most of these regulations of the use of AI in administrative law settings will mandate, or make a reference to, some kind of cost–benefit analysis of the social utility of the deployment and use of the system, in the way described by Sunstein.[19] The test to start employing automated systems in this context is one that compares an existing procedure in which humans participate against the efficiency, savings, reliability, risks of mistakes and harms, and other social and cultural aspects of the automated systems.

Probably the only real complication from a regulatory point of view for these systems is the decision to shift from one model to another. I have considered this problem with Vijeyarasa in relation to the VioGén, a computer-based system used for the assessment of the level risk of revictimisation of victims of gender-based violence in Spain.[20] If an AI-based system is considered to be ready to deliver an output better than a human qualitative assessment or one based on traditional statistics, what is the degree of outperformance compared to humans, or the level of reassurance necessary to make that shift, and how much capacity should be left to bureaucrats to override the system's decisions? These are not easy questions, but they are not difficult to visualise: should the standard for accepting automation be performing better than an average bureaucrat? Better than the bureaucrats with the best track records? Or when the risk of expected errors is considered as reduced as possible? At similar levels of performance, should cost be considered?

These are decisions that administrative law could explicitly leave to the discretion of bureaucrats, establish *ex ante* binding rules or principles, or leave it to the judiciary to consider it if a complaint is made. Again, not easy questions, but decisions that could be addressed within the principles that we are familiar with. In the end, the reasoning is not that dissimilar from a decision to externalise to a private provider a service hitherto delivered by the state.

[18] BG Peters, *Politics of Bureaucracy*, 5th ed (Routledge, 2002) 35.
[19] CR Sunstein, *The Cost–Benefit Revolution* (MIT Press, 2019).
[20] J-M Bello y Villarino and R Vijeyarasa, 'International Human Rights, Artificial Intelligence, and the Challenge for the Pondering State: Time to Regulate?' (2022) 40 *Nordic Journal of Human Rights* 194–215 at 208–9.

To be clear, I am not suggesting that there is anything intrinsically wrong with focusing our (regulatory) attention on these issues. I believe, however, that this view encompasses a very narrow understanding of what AI systems could do in the public sector and the legal problems it can create. This approach is conceptualised in terms of efficiency and the hope that AI can finally deliver the (so far) unmet promise of the productivity revolution that was expected from the massive incorporation of computers in the public officials' desks.[21]

From that perspective AI could be a key element of *that* Automated State. AI systems could be optimised to limit the variance between decisions with similar or equal relevant attributes. Consequently, AI-driven systems could be the best way to reach a reasonable level of impartiality, while fulfilling mundane tasks previously performed by humans.

Obviously, this cannot happen without maintaining or improving the rights of those individually or collectively affected by these automated decisions. Administrative law would need to ensure that the possible mistakes of these 'approved and certified' systems can be redressed. The legal system must allow affected parties to challenge outputs that they believe do not correctly implement policy. This could be, at least, on the basis of a possible violation of any relevant laws for that policy or a lack of coherence with its objectives, or with other relevant rights of the person or entity affected by the output of the system.

Therefore, the only need for reforms (if any) for administrative law in *this* Automated State is to (i) create a path to pre-validate the system; (ii) create guidance or determine when to change to such a system; and (iii) enable parties affected by its outputs to complain and challenge these decisions.

Other chapters in this book look at this third point in more detail, but I see it as requiring affected parties to go 'deeper' into the automated (or machine-supported) decision. The affected party, alone or in conjunction with others affected by the same or similar decisions from that system, need to be able to – at least – (i) explore why their decision can be distinguished from similar cases deserving a different administrative response; (ii) be able to raise new distinguishing factors (attributes) not considered by the system; and (iii) challenge the whole decision system on the basis of the process of pre-certification of the system and its subsequent monitoring as the system learns.

Generally speaking, the type of legislative reform necessary to accommodate this change will not create excessive friction with the approaches to administrative law

[21] There is a societal expectation that AI-driven systems can materialise the productivity jump that computers did not bring, and respond to Nobel Prize laureate Robert Solow's quip that 'you can see the computer age everywhere but in the productivity statistics'. 'Why a Dawn of Technological Optimism Is Breaking' (16 January 2021) *The Economist*; 'Paradox Lost' (11 September 2003) *The Economist*.

https://doi.org/10.1017/9781009334297.011 Published online by Cambridge University Press

already in place in civil and common law systems.[22] Essentially, the only particularity is to be sure that the rights of the parties affected by administrative decisions do not get diluted because the administrative decision comes from a machine. The right to receive a reply, or to an intelligible explanation, or to appeal a decision considered illegal should be adapted, but not substantially changed.

Perhaps the concept of the 'organ' in civil law systems and the allocation of responsibility to the organ, which in practice makes administrative law a distinct area of law, with a different logic from the civil/criminal dichotomy still dominating the common law system,[23] could make the transition easier in civil law systems. The organ, not the bureaucrats or their service, is responsible for its outputs. However, certain rules about the burden of proof and the deference towards the state in continental systems could make it more difficult to interrogate the decision-making process of a machine.

Finally, in terms of administrative law, it is even possible to envisage a machine-driven layer of supervision or control that could monitor human action, that is, using AI to supervise the activity of public officials. One could imagine a machine-learning system which could continuously check administrative outputs created by human bureaucrats alerting affected parties and/or bureaucrats when it detects decisions that do not appear to align with previous practice or with the application of the normative and legal framework. Such an Automated State could even increase the homogeneity and predictability of administrative procedures and their alignment with the regulatory regime,[24] therefore increasing trust in the public system.

In this scenario, the Automated State will not (for the time being) replace humans, but work alongside them and only reveal itself when there is a disparity of criteria between the output of the human bureaucrat and the automatic one. The existence of this Automated State cohabiting with a manual one may require different administrative rules for human-made decisions. When decisions diverge, possible options may involve an obligation to notify affected parties of this divergence and, perhaps, granting them an automatic appeal to other administrative entities, or requiring reconsideration by the decision-maker, or imposing on the human decision-maker an obligation of more detailed and explicit motivations. In this state of automation, the human administrative decision will not be fully acceptable unless it aligns with the expected one from the Automated State. And, yet, we can still address these situations with a *lex specialis* for the automated decision, remaining within the logic and mechanisms of 'traditional' administrative law.

[22] Ombudsman New South Wales, *The New Machinery of Government: Using Machine Technology in Administrative Decision-Making* (Report, 2021).

[23] JAS Pastor, 'La teoría del órgano en el Derecho Administrativo' (1984) *Revista española de derecho administrativo* 43–86.

[24] Cordella and Tempini, 'E-Government and Organizational Change', 280.

Having now covered the easier of the two transitions, it is now the time to consider the other Automated State, the one that liberal-democratic legal systems could find most difficult to accommodate. The tale of the Automated State that designs or evaluates policy decisions.

7.3 REGULATING THE UNSEEN AUTOMATED STATE

As noted in the Introduction, AI can be harnessed by public authorities in ways that have not been seen before. The idea of a digital twin, for example, alters the logic behind the discretion in the decision of public authorities, as it makes possible to envisage both states of a world, with and without a decision.

If we take another step in the same direction, one could even assume that in the future the design and establishment of policy itself could be delegated to machines (cyber-delegation).[25] In this scenario, AI systems could be monitoring opportunities among existing data to suggest new policies or the modification of existing regulations in order to achieve certain objectives as defined by humans or other AI systems.

Yet, for the purpose of this chapter, we will remain at the level of the foreseeable future and only consider systems that may contribute to policy determination. The discussion below also assumes that the systems are correctly designed and operate as they are expected.

This type of automation of the state involves expert systems that are considered to provide higher levels of confidence about choices in the policy-making process. This view of the Automated State sees AI systems as engineered mechanisms 'that generate[...] outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives', in line with current thinking in the global standardisation process.[26]

This corresponds to existing observations in governance theory that note that 'the transfer of governmental decision-making authority to outside actors occurs along a continuum'.[27] A public authority generally decides on policy through an output generated by one of its employees (elected or appointed) or a committee of them. How to reach that policy decision could be left to the employees of that public authority, reached through a system of consultation, or fully deferred to a committee of experts.

Regardless of how the decision is reached, the essential element is that the decision process is oriented towards the achievement of an implicit or explicit set of human-defined objectives. Achieving these objectives is the raison d'être of the policy decision, even if, from a social point of view, the ultimate motivation, and,

[25] Gulson and Bello y Villarino, 'AI in Education'.

[26] 'ISO/IEC 22989:2022(en)' (2022) sec. 3.1.4.

[27] M Shapiro, 'Administrative Law Unbounded: Reflections on Government and Governance Symposium: Globalization, Accountability, and the Future of Administrative Law' (2000) 8 *Indiana Journal of Global Legal Studies* 369–78 at 371–72.

therefore, the legitimacy of the decision to set these precise objectives, may have been spurious (e.g., to unjustifiably favour a certain service provider over others). If the advice to the decision-maker is assisted by an AI system, however, that objective needs to be explicit as it is what the system will try to achieve and optimise in relation to other factors.

Allow me, however, to explain the consequences of this statement, before exploring these objectives. The state, as an agent, does not act on its own behalf. The existence of the modern liberal state is based on the founding principle that it does not act on its own interest, but as a human creation for the benefit of its society. The human-defined objectives are the reason for its existence, the state being a tool to achieve them.

Leaving aside if this is actually the case – diverging from those who see the state as better described as a mechanism for preservation of certain parts of that society or more theoretical discussions about the role of the state – in this section I assume that decision-makers are honest about those objectives or boundary conditions.[28] As noted in the previous sections, what matters for systems that merely apply policy to reach outputs is to correctly reflect that policy in those outputs. Broader objectives such as fairness through redistribution, or equality of opportunities must be embedded in the policy design, the outputs just being the automated application of that policy. Here, the policy is what is being created by the Automated State, so the system will design or propose a policy that optimises those objectives.

In societies that democratically elect its decision-makers, one can assume that some of these objectives can come from different sources, such as:

1. Those determined by basic legal norms that constrain the action of public authorities. This is the case, for example, of constitutional rules, such as 'no discrimination on grounds of age or socioeconomic grounds', or a mandate to redress inequality derived from socioeconomic grounds or a 'right to access a no-fee system of quality education until the age of 16'.[29]
2. Those determined by the objectives hierarchically established at higher levels of decision-making. For example, one could consider the programme from a central government, or the priorities established at the ministerial level – and the principles explicated therein – as a restriction to the action of lower hierarchical levels, especially when materialised in formal directives. For example, in the fiscal context, one objective could

[28] That is, not cheating the process, for example, through entering into the automated system a series of acceptable objectives until they reach a desired output for other reasons, that is, their real hidden objectives.

[29] For a sample of countries having the right of education in their constitutions, see S Edwards and AG Marin, *Constitutional Rights and Education: An International Comparative Study* (2014).

https://doi.org/10.1017/9781009334297.011 Published online by Cambridge University Press

be increasing the fight against fraud or, in the education context, improving the standardised results of students from disadvantage backgrounds.

3. Those that are determined by the specific decision-maker (organ or individual), who is formally in charge of making that decision. For example, in the tax context it could be accepting that more exhaustive detection of fraud would be at the cost of more administrative complaints from honest taxpayers that would be incorrectly identified. In the education context it could be a limit in the amount of resources that could be allocated to improving educational standards overall.

In all three cases the objective is the key element for the development of policy. An Automated State in which AI systems are designed to optimise these objectives will, in principle, derive its legitimacy and legality from these objectives. More importantly, the sequence of objectives listed above can be seen as hierarchical, with policymakers assisted by these AI systems bound by the objectives established by the superior levels. As an example, a decision-maker on the lowest level of hierarchy who sets the level of expenditure at this lowest level (district, local council, federated state, or national level) for public (government paid) education could not accept any recommendations from the Automated State that could suggest as optimal interventions those expected to deliver a significant improvement for overall academic standards for 99.9 per cent of the students of that administrative level, but would not offer free education for the 0.1 per cent living in the most remote communities if there is a constitutional mandate to offer free education for all. A proposal that would involve the exclusion of even one person would not be acceptable. Similarly, an option that improves the academic results for all at a given cost, but forces students from the most deprived backgrounds to separate from their families would be a violation of a tier 1 objective, and, therefore, not acceptable either. A correct design of the AI system producing the recommendation should not even generate these options.

Obviously, not all objectives follow this neat hierarchical structure. Sometimes the systems could offer recommendations for policy options that are seen as trade-offs between objectives at the same level. Some other times, there could be enough flexibility in the language of the boundary conditions that, at least formally speaking, it would not require to build those boundary conditions into the system. This would allow systems to generate some proposals that would not be accepted under a stricter objective or a different reading of the wording of the objective.

For example, a system may be allowed by humans to suggest an education policy that is expected to achieve a significant improvement for 99.9 per cent of the students. In this case policymakers tasked with creating a policy to improve standardised scores may decide to allow systems to consider this option, if they knew that they could meet the formal requirement of providing free education for all students through other means or policies. That could for example involve providing untested

remote self-learning program to students for free. This would be feasible in policy settings where the boundary condition is just 'providing non-fee education' without qualification of '(proved) quality'.

As we know, it is not unusual for general mandates to be unqualified, particularly at the constitutional level, and see the qualifications being derived by interpretation from other sources (human rights principles or meaningful interpretations from high-ranking courts). In any case, it is how humans decide to translate those mandates into the system objectives what matters here.

Yet, this kind of problem may still not be that different from what systems of administrative control are facing today. The level of discretion is still added into the systems by humans and this concrete human choice (the decision to place other options within the scope of analysis) is still the one that could be controlled by courts, Ombudsman, or any other systems of administrative checks.

A second type of problem appears when the system is showing that certain options are superior to others, but benefit some groups of people differently. For example, a system that is expected to improve the results of all students, but improve the results of students from advantaged socioeconomic backgrounds by 10 per cent and those from disadvantaged backgrounds by the same 10 per cent would not be generated as a recommendation by a system which is requested to produce only options that are also expected to redress inequality. However, the same system could recommend the next best option at the same cost, which is expected to improve the results of the first group by 7 per cent and the second by 9 per cent, as this option does address inequality, which was a requirement set by humans to the system.

Favouring the latter proposal may seem absurd from a (human) rational point of view. The first suggestion is clearly superior as it would see all students being better off overall in terms of academic performance. Yet, only the second system would meet the objectives manifested in boundary conditions. A correctly built system would respect the hierarchy of objectives. Given that redressing inequality is more likely to be a constitutional or general mandate and, therefore, trump improving results – which is more likely to be an objective set at a lower hierarchical level – the first option would never be offered as a suggestion to the policymaker.

In this case, a better approach would be to allow the Automated State to present the first option to policymakers as far as the expected outputs are clear and the violation of the boundary condition is explicit. This would allow policymakers to simultaneously intervene in other ways to redress inequality. AI systems do not live in a policy vacuum, so it is important to design them and use them in a way that allows for a broader human perspective.

A third type of problem could occur when the system is designed with an added level of complexity, presenting the options in terms of trade-offs between different objectives at the same level.[30] For example, the choice could be offered to the

[30] Gradient Institute, *Practical Challenges for Ethical AI* (Report, 2019) 8.

decision-maker as policies that are expected to deliver overall improvements of educational standards, for all students, with a bigger gain for those from disadvantaged backgrounds (i.e., meeting all the boundary conditions and objectives), but expressed in terms of cost (in monetary units) and levels of overall improvement. Then it would be up to the decision-maker to decide which option of the many possible ones would be preferred. In this case, the main problem is one of allocation of resources, so this could initially be left to human discretion. However, as public resources are limited, if different AI systems are used to automate policy-making, setting a limit for one of these trade-offs would affect the level of trade-offs for other recommendation systems operating in other policy areas.

This could be intuitively grasped in the tax context. Imagine a public authority tasked with maximising tax revenue at the lowest cost within the legal boundaries. The system assessing anti-fraud policy may recommend an optimal level of investment in anti-fraud and establish the identified taxpayers that should be checked. Other system may be used to recommend possible media campaigns promoting compliance. This other system may suggest an optimal level of investment and the type of campaigns expected to give the highest return. Yet, it is possible that the level of resources available may not be enough to follow both suggestions. A broader system could be created to optimise both systems considered together, but what could not be done is considering each of the systems in isolation.

Looking together at these three types of problems gives us an idea about how this Automated State is different. For the systems discussed in Section 7.2, those that replace humans, I indicated that the most promising regulatory approach is the one that focuses on the systems and the testing beforehand and then shifts to monitoring of the outputs. As the bulk of the effects of each automated decision will be centred around a limited (even if large) number of individuals, the affected parties will have an incentive to raise their concerns about these decisions. This could allow for a human (administrative or judicial) review of these decisions according to the applicable rules. The automated outputs could be compared with what humans could do, according to the applicable administrative law, in those circumstances. This process would confirm or modify the automated decision and the automated systems could be refined to learn from any identified errors.

However, for the systems discussed in Section 7.3, that are used to do things that humans cannot do, especially in terms of policy design or supervision, it is impossible to proceed in such a way. Any challenge of a concrete decision could not be compared with what a human could do. Any disagreement about the reliability of the system would be too complex to disentangle.

Yet, there are aspects of the process that would still need to meet societal standards about adequate use of resources, fulfilment of superior principles of the state, or, more generally, the need to meet the state's positive obligations to protect human rights, remove inequalities, and redress violations of rights of individuals or groups.

At the very least there are three elements regarding how humans interact with the systems that generate the outputs that could be considered.

First, humans must test the systems. To grant some legal value to the recommendations of these systems – for example, to demand more from policymakers that deviate from their recommendations – this type of Automated State must be tested in real-life, real-time conditions. In the next section I explain in more detail what I mean by this point. Suffice to note here that systems tested only against data from the past may not perform well in the future and their legal usefulness as a standard for the behaviour of policymakers may therefore be undermined.

Second, humans must set the objectives that the system is meant to optimise (and suggest ways to achieve) and the boundaries that the suggestions are not meant to trespass. Which objectives and boundaries are incorporated into the system and how they are hierarchically placed and balanced can be explained and the legality of those choices controlled.

Third, humans must translate automated suggestions into policy. The example of the AI system used for assessing quality of teaching in the United States discussed in Chapter 10 of this book[31] is a perfect example of this point. Even if we trusted that the system was correctly evaluating the value of a teacher in terms of improvement of the results of their students, the consequence attached to those findings is what really matters in the legal sphere. Policymakers using such a system to assess quality of teaching could decide to fire the lowest performing teachers – as it was the case in Houston – or to invest more in the training of those teachers.

7.4 PREPARING FOR THE TWO TALES OF THE AUTOMATED STATE

In the previous sections, I discussed the two different tales of the Automated State and the distinct legal implications that each tale involves. This, however, was an oversimplification. Going back to the VioGén system presented above, one can today see a system of implementation, typical of the first tale of the Automated State, as it assesses each individual woman based on their risk of revictimisation. The suggested assessment, if accepted by the human decision-maker, automatically triggers for that victim the implementation of the protection protocol linked to her level of risk. Yet, VioGén could easily become a policy design tool. For example, it could be repurposed to collate all data for all victims and redeveloped into a system that allocates resources between women (e.g., levels of police surveillance, allocation of housing, allocations of educational programmes, suggestions about levels of monitoring of restraining orders for those charged with gender-based violence). If we consider every automated system a potential policy tool, we may be moving towards an excessive degree of administrative control of policy-making. As policymakers will

[31] *Houston Federation of Teachers Local 2415 et al v Houston Independent School District*, 251 F. Supp. 3d 1168 (2017).

have much more and richer data, administrative law could be used to question virtually any policy decision.

At the other extreme, one could think that it could be better to revert to almost complete deference to the discretion of policymakers. If we think of policy-making as a black box driven by criteria of opportunity or the preferences of high-ranking elected officials it is difficult to justify the need for a new type of administrative law for these situations, even if the policymakers are better placed to assess the consequences of their decisions. One can, for example, imagine the decision of a public authority to approve a new urban planning policy after a number of houses are destroyed by floods. The new policy may be so different to previous practice that its effects in case of another flood cannot be assessed by an AI-driven recommendation system. The system, however, can suggest several minor modifications that are expected to be enough to avoid a repetition of the situation. In this case the ultimate purpose of the new policy may be to increase resilience of the housing in case of new floods, but the real value of the initiative is to convey that public authorities are seen as reacting to social needs.

The expected evolution of the first type of the Automated State could also support deferring to the discretion of policymakers and ignoring the new tools of the Automated State from an administrative law perspective. As more decision-making is automated at the level of implementation, a reduction of variance should be expected. The effect in the world of these outputs could then be analysed in real time and the outputs will speak for the policies they implement. Public office holders would then be accountable if they fail to modify policies that are generating undesirable outputs. The effects of a change in policy that is implemented through fully automated means will be the basis to judge that policy. Policy design will not only refer to 'design', but also the choice and design of the automated tools that implement it.

In my view, none of these options are reasonable, so it is necessary to start developing new principles that acknowledge the legal relevance of these new tools in policy-making, without separating ourselves excessively from the process. The absolute deference to policymakers choices, even if tempting, would be a reversal of the positive 'erosion of the boundaries separating what lies inside a government and its administration and what lies outside them' or, in other words, of the transition from 'government' to 'governance'.[32]

A way to illustrate this latter point would be to consider the French example, and its evolution from a black-box State to an *administré*-centred one.[33] This transition, induced – according to a leading French scholar – by Scandinavian-, German-, and EU-driven influences, has forced administrative law to go beyond traditional rights

[32] Shapiro, 'Administrative Law Unbounded', 369.

[33] P Gérard, 'L'administré dans ses rapports avec l'État' (2018) 168 *Revue française d'administration publique* 913–23.

in French law (to an intelligible explanation, to receive a reply, to appeal a decision considered illegal) into a regime where the *administré* can be involved in the decision-making process and is empowered vis-à-vis the State.[34] It is not just the output, but also the logic behind the process that matters.

If the reasons for policy decisions matter, how can we then use the Automated State to demand better accountability for those decisions? Trusting this Automated State blindly or inextricably binding decision-makers to its decisions does not appear to be a good option, even if we have tested the AI systems according to the most stringent requirements. My suggestion is to develop a few principles or heuristics that could guide us in the process of reform of administrative law.

The first – and most essential from my point of view in a technology without historic track record of performance – is that systems designed to make predictions about impacts of public actions in the future need to have been tested in real conditions. This Automated State could only be relied upon for the purpose of legal assessments of policy decisions, if the predictions or suggestions of its systems have been proven to be reliable over a given number of years before the date of the decision.

Systems that are 'refined' and reliable when tested against the past cannot be a legal basis to contest policy decisions. Only real-life experiments for policy design without 'the benefit of hindsight' should matter. In these cases, deference should be paid to policymakers to the same degree as before. However, for learning and testing processes an adequate record of use should be kept – that is, systematically recording how the system was used (for testing purposes) in real-time conditions.

Secondly, we should be flexible about setting boundaries and objectives. Administrative rules should not impose designs that are excessively strict in terms of hierarchy of objectives, as some of the objectives can be addressed by different policies at the same time, not all covered by the automated systems. For those cases, the systems should be designed to allow for the relaxation of the boundary conditions (objectives) in a transparent manner, so policymakers can assess the need for other interventions. In the example above about the education systems, a rigid translation of legal principles into data could blindside us to policy options that could be adapted further to respect legal boundaries or even be the reason to adapt those boundaries.

Finally, decisions that deviate from those suggested by legally reliable automated systems should be (i) motivated by decision-makers in more detail than traditionally required; and (ii) the selected (non-recommended/Pareto-inferior) policy should be also assessed with the relevant AI systems before implementation. The results of that assessment, the policymaker motivation, and all connected information should be made – in normal circumstances – publicly available. This would allow the improvement of systems, if necessary (e.g., incorporating other considerations),

[34] Ibid.

and allow better administrative or judicial control of the decision in the future. Guidance could be extracted from decisions that override recommendations of environmental impact assessment, where an administrative culture that relied on discretion rather than law – for example in the English context[35] – has traditionally been an obstacle to the effective judicial control of those decisions. Discretion should be accepted as an option as far as it is explicitly justified and, hopefully, used for developing better automated systems.

[35] J Alder, 'Environmental Impact Assessment – The Inadequacies of English Law' (1993) 5 *Journal of Environmental Law* 203–20 at 203.

https://doi.org/10.1017/9781009334297.011 Published online by Cambridge University Press