Regulating the Robot: A Toolkit for Public Sector Automated Decision-Making

Benjamin Cartwright*

Abstract—The use of automated decision-making (ADM) technologies by public sector bodies has developed significantly in recent years. As a result, decisions are increasingly made with limited, or no, human involvement. The need to regulate such technologies is therefore imperative. This article responds to the recent proposal of Lord Sales to establish an 'Algorithm Commission', to regulate and advise public bodies on ADM system design and scrutiny. While such a proposal is worthy, it is argued that a better method by which to regulate ADM technologies is through a toolkit of regulatory mechanisms, by deploying targeted impact assessments, codes of practice, technical tools and the existing data protection framework.

* BCL candidate, Harris Manchester College, University of Oxford.
This paper was written during my LLB at University College London. I am grateful for the insightful guidance and comments on earlier drafts by Professor Rick Rawlings, Chris Moss and the OUULJ Editorial Board. All errors remain my own.

Introduction

'The algorithms of the law must keep pace with new and emerging technologies'.[1] Nowhere is this truer than in the context of public sector decision-making, where the recent uptake of artificial intelligence (AI) based automated decision-making (ADM) systems has been astonishing. From use in tax audit, to improving health standards and policing, automated processes have served to transform the administrative state, using technology to raise standards, reduce error rates and cut costs. A recent US study indicates that ADM systems are now used in more than 45% of federal agencies, with many more such systems in the pipeline.[2] This increasing ubiquity, and the challenges associated with ADM systems, makes it essential to effectively regulate them: their deployment will only continue in the coming years.

Considering this, Lord Sales has proposed an 'Algorithm Commission' to regulate and advise Parliament, the courts and government, on matters of ADM system design, scrutiny and adjudication.[3] In this article, it will be argued that while such a Commission may be helpful in ensuring accountability, regulatory *bodies* such as that proposed by Lord Sales are an inferior method of regulating ADM systems, when compared to instrumental legal and regulatory *mechanisms*, such as algorithmic impact

---

[1] *R (Bridges) v Chief Constable of South Wales Police* [2019] EWHC 2341 (Admin) [2020] 1 WLR 672 [1].

[2] David Freeman Engstrom, Daniel E Ho, Catherine M Sharkey and Mariano-Florentino Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies* (Administrative Conference of the United States 2020) 6.

[3] Lord Sales, 'Algorithms, Artificial Intelligence and the Law' (Sir Henry Brooke Lecture for BAILII, 12 November 2019) < https://www.supremecourt.uk/docs/speech-191112.pdf> accessed 10 January 2020.

assessments, codes of practice, technical tools and the existing data protection framework. It is argued that a new legal toolkit, fashioned around such regulatory mechanisms, will be better placed to mitigate the accountability risks associated with ADM system development. Ultimately, intrinsic regulation of ADM systems, albeit necessarily supported by a regulatory body, is more likely to guarantee adherence to the overarching legal regime. It is therefore not argued that there should be *no* role for a regulatory body; *a* body will be needed to deploy the toolkit argued for, and to ensure that the instrumental mechanisms meet their intended purposes in ensuring algorithmic accountability. As governance is increasingly digitised, the need to modernise accountability mechanisms which public bodies may deploy only grows.

What follows takes a three-part structure. Part 1 conceptualises and briefly surveys various legal issues related to ADM. Part 2 focuses on the ambit of an algorithm regulator responsible for ensuring *ex-ante* accountability. Part 3 outlines the proposed legal toolkit which the regulator would be responsible for deploying and overseeing. The focus of the regulatory toolkit proposed will be on *ex-ante* regulation of ADM systems, given that hitherto, most discussion of ADM system regulation has been to ensure *ex-post* (judicial) accountability. This said, it will be argued that if the proposed toolkit is adopted, both *ex-ante* and *ex-post* accountability will be improved.

1. Conceptualising Automated Decision-Making

A. Terminology

Automated decision-making refers to a process which is, partly or wholly, automated by a computer system, with the aim of increased efficiency and volume of decision-making.[4] The term algorithm generally has a wider meaning, constituting a logical process in which a series of rules or conditions are followed to meet defined ends.[5] Essentially, if conditions A, B and C are met, X will occur as a result. Since all ADM systems are founded on algorithms, the terms in this article are used interchangeably.

Artificial Intelligence takes algorithmic processes a step further, using technological automation to deploy them either somewhat or entirely automatically. The tasks completed by AI can be more complex than more analogue algorithmic processes, which previously depended on human judgment.[6] A *machine learning algorithm* (MLA) takes this sophistication and complexity a step further still, identifying patterns in vast data sets, rendering them into a manner usable for other purposes such as policy delivery.[7] While most ADM systems remain as relatively typical algorithms, some constitute incredibly sophisticated MLAs. Such systems have been—and are increasingly being—deployed in service of the aims of the public sector; this may range from the automation of simple tasks (a basic example might be the

[4] Michael Veale and Irina Brass, 'Administration by Algorithm? Public Management Meets Public Sector Machine Learning' in Karen Yeung and Martin Lodge (eds) *Algorithmic Regulation* (OUP 2019) 123.
[5] Adapted from Sales (n 3) 2.
[6] ibid 4.
[7] Veale and Brass (n 4) 121.

development of e-filing systems to save time and complexities of paper filing), to extremely complex decision-making structures, which can entirely remove the need for human intervention.

B. Regulatory Challenges

There are undoubtedly benefits to the employment of ADM systems, not least their use in high-volume areas of government; these can boost government capacity and improve service delivery, for instance in fraud detection, police resource allocation and the risk-scoring of prisoners.[8] However, they are not without their challenges, four of which will be discussed: these are I. legality; II. opacity; III. bias; and IV. contractor accountability. Overcoming these challenges is the central purpose of ADM system regulation.

I. Legality

It is essential that any public sector algorithm accurately represents the underlying legal framework; however, when we go beyond 'highly specific… semantically un-troubling domains',[9] to complex administrative service delivery, accurately representing entire statutory schemes can be challenging. This is particularly so given the value-laden nature of law: concepts like fairness and justice surely cannot be reduced to a few lines of code; they are not quantifiable concepts, they represent value judgments which (human) judges often struggle to grapple with, and can give rise to divergent interpretations.[10] Nor would it necessarily be a good

[8] ibid 137.
[9] Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" is Probably Not the Remedy You Are Looking For' (2017) 16(1) Duke Law and Technology Review 18, 24.
[10] As Brad Smith and Carol Ann Browne have written, '[h]ow can the world converge on a singular approach to ethics for computers when it

thing to algotithmise such concepts: one of the benefits of such concepts is the ability to deploy them with a degree of flexibility, to meet the justice of the case. The idea of quantifying and weighting fundamentally incommensurable factors, while something that lawyers may be accustomed to, is anathema to most coders.[11] Further, an ADM system must continue to appropriately perform its designated functions at the operation stage. A particular risk is that once deployed, an ADM system starts to change the way in which a decision is carried out (e.g. as a result of machine learning), in a manner imperceptible to the public body or the decision subject.[12] To avoid this, it is important that human agency is maintained over the MLA to some degree and that independent (human) judgment of policymakers, who have experience making such decisions, is not undermined.[13] While it is true that in ADM system design, there will be human involvement to ensure that the system adheres to the statutory and policy framework (since at present, most ADM systems depend on human coders to set their parameters), ensuring that this continues once the code has been written, and the system has been deployed, remains crucial.

---

cannot agree on philosophical issues for people?', *Tools and Weapons: The Promise and the Peril of the Digital Age* (Penguin 2019) 207.

[11] Veale and Brass (n 4) 133. It is also important to note that in addition to coders' inability to measure incommensurable factors, computers themselves may not be able to (at present) replicate the flexibility that humans have in weighing up such factors.

[12] Marion Oswald, 'Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power' (2018) 376 Philosophical Transactions of the Royal Society A 20170359, 14.

[13] Karen Yeung, 'Algorithmic Regulation: A Critical Interrogation' (2018) Regulation & Governance 505, 516.

II. Opacity

The legitimacy of ADM systems risks being undermined given their lack of transparency; such opacity can be intentional or inadvertent. Intentional opacity may result from national security or public policy considerations, and a consequent refusal to disclose the relevant code. For example, it could be dangerous to allow potential terrorists to dissect an ADM system which is used to identify potential flight risks, or to allow potential tax cheats understand the process by which tax records will be selected for audit.[14] Intentional opacity may also arise through an unwillingness of private firms to disclose the system code. For instance, Apple refused to assist the FBI to gain entry into the iPhone of a convicted terrorist, for fear of the chilling effect it may have on private sector innovation.[15]

More challenging still is inadvertent opacity. If millions of variables affect a decision, it may be unclear how any single variable ultimately affects the system. In an ADM system deployed by Durham Constabulary, which supports custody officer decision-making, there are 4.2 million decision-points, each dependent on previous data.[16] Hence, even if the inputs

---

[14] Adapted from Joshua A Kroll, Joanna Huey, Solon Barocas, Edward W Felten, Joel R Reidenberg, David G Robinson and Harlan Yu, 'Accountable Algorithms' (2017) 165 Pennsylvania LR 633, 638-639. See also Veale and Brass (n 4) 138.

[15] Arjun Kharpal, 'Order to hack iPhone for FBI "chilling": Tim Cook' (*CNBC*, 17 February 2016) <https://www.cnbc.com/2016/02/17/apple-order-to-hack-iphone-for-fbi-in-san-bernardino-case-chilling-tim-cook.html> accessed 20 February 2020. See also Apple, 'A Message to Our Customers' (16 February 2016) <https://www.apple.com/customer-letter/> accessed 11 March 2021.

[16] Marion Oswald, Jamie Grace, Sheena Urwin and Geoffrey Barnes, 'Algorithmic risk assessment policing models: lessons from the

behind ADM systems are easily understandable, given the 'black-box' nature of the processes and the contextual application of the algorithms, outputs are not always predictable.[17] Given that many ADM systems 'learn' from contextual data and change their application accordingly,[18] mere inputs will never be sufficient to understand the behaviour of the ADM system; comprehensive algorithmic audit, taking into account source data *and* subsequent contextual data is important. This is difficult to achieve prior to deployment.[19]

Therefore, the solution to the problem of opacity cannot simply be transparency through disclosure of the source code. Disclosure will not assist understanding of algorithms, which can be notoriously complex, particularly when we remove the source code from the contextual framework in which the system operates.[20] This is as much the case for administrators themselves, as much as it is for members of the public; the inherent complexity of digital governance, and the fact that most civil servants are not programming experts, mean that they will be unlikely to appreciate fully the tools they are using, and the impact that the ADM systems will have. This is not like analogue policy design and implementation, which one would expect to be more intuitively understandable than complex ADM system design. Hence, regulation of ADM systems is important not only to

---

Durham HART model and "Experimental" proportionality' (2018) 27(2) Information & Communications Technology Law 223.

[17] Marijn Janssen and George Kuk, 'The challenges and limits of big data algorithms in technocratic governance' (2016) 33 Government Information Quarterly 371, 374.

[18] See discussion of MLA in the first part of this paper.

[19] However, the risks can be mitigated by way of adoption of the methods proposed in the third section of this paper.

[20] Anton Vedder and Laurens Naudts, 'Accountability for the use of algorithms in a big data environment' (2017) International Review of Law, Computers & Technology 206, 215.

ensure the accountability of public sector decision-making to the public, but also to ensure that the ADM systems remain understandable and scrutable tools to *serve*, rather than potentially undermine, policy delivery. As shall be explained in the third section of this paper, technical tools in particular would greatly assist such an understanding of ADM systems.

Even if algorithms are published, there is a risk of their purpose being undermined. This is possible in two senses, as Veale and Brass outline: first, it may permit the targets of ADM systems—such as possible terrorists, in the context of terror risk detection system—to interrogate the methods and datapoints used, so as to 'play the system' and escape scrutiny.[21] If ADM systems are indeed exploited by the very people they are intended to regulate, the effect any risks which do materialise could be catastrophic. This is not to suggest that *human* decision-making processes are not themselves flawed and subject to manipulation; it is just to emphasise the inherent risks of publication of the algorithmic code, which can go beyond mere unintelligibility.

Second, publication of an ADM system's datapoints may undermine public confidence in the system, particularly if there is a coded-in risk of false negative results.[22] This second point is worth underlining; even if we accept the capacity of humans to make mistakes, 'it seems somehow less palatable that a bureaucracy has decided to deploy a system with a [particular]… rate of false negatives.'[23] While such a rate of false negatives may well be an improvement over the (often flawed and opaque) decision-making processes of humans, and bring consequent benefits to stakeholders, it is the appearance of *intended* failure that the public may find particularly galling. Even with strong

[21] Veale and Brass (n 4) 138.
[22] ibid.
[23] ibid.

contextual information demonstrating an improvement in decision-making capacity, trust in public bodies to do the right thing may decline.[24]

III. Bias

A third problem is that of machine bias. In administrative law, a system will show apparent bias, if 'the fair-minded and informed observer, having considered the facts, would conclude that there was a real possibility of bias'.[25] Hence, an ADM system could be biased if it has an internal model which does not produce fair and consistent outputs, in the sense that it consistently benefits or disadvantages a given group; in such a case, apparent bias will likely exist.[26] Whereas prima facie, ADM systems may *benefit* from the absence of active or passive human bias,[27] it is well-established that any code may be value-laden, perhaps (un)intentionally reflecting the coder's biases. One study, by Kleinberg,

---

[24] A more limited measure of disclosure may well be possible, and may go a long way in accounting for the problem of opacity. However, such disclosure would still have to contend with practical questions of public policy and national security; suitable safeguards such as those that are recommended in the third part of this paper should be adopted, to mitigate such risks.

[25] *Porter v Magill* [2001] UKHL 67, [2002] 2 AC 357 [103] (Lord Hope). The more complex the issue, the more 'informed' that the fair-minded observer will have to be; this will often result in the test being no more than the actual view of the court being reached: *Hart v Relentless Records* [2002] EWHC 1984, [2003] FSR 36. Further discussion of the test is beyond the scope of this paper; for further information, see Lord Woolf, Jeffrey Jowell, Andrew Le Sueur, Ivan Hare and Catherine Donnelly, *De Smith's Judicial Review* (8th edn, Sweet & Maxwell 2020) 10-012 ff.

[26] Jennifer Cobbe, 'Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making' (2019) 39 Legal Studies 636, 654.

[27] Veale and Brass (n 4) 126.

Mullainathan and Raghavan, indicated that elimination of bias may never be possible.[28] Further, any mistake in the coding, or any widespread social bias reflected in training data used to set up MLAs, may result in biased or discriminatory decisions, which may be difficult to identify given the aforementioned impenetrability of algorithms.[29] This risk of entrenching bias is highlighted by a recent threat of judicial review of a Home Office ADM tool which would filter visa applications; the applicants' argument, that the ADM system discriminated on the basis of race, was enough to persuade the Home Office to 'discontinue' and 'redesign' the algorithm. In redesigning the algorithm, the Home Office will consider in particular 'issues around unconscious bias and the use of nationality generally in the streaming tool.'[30]

IV. Contractor Accountability

Fourth, the development of ADM systems has involved wide-ranging public contracting and involvement of the private sector; this is effectively a continuation of the use of private contractors

[28] Jon Kleinberg, Sendhil Mullainathan and Manish Raghavan, 'Inherent Trade-Offs in the Fair Determination of Risk Scores' (2016) <https://arxiv.org/abs/1609.05807> accessed 27 December 2020 (cited in Cobbe (n 26) 654).

[29] Andrew Tutt, 'An FDA for Algorithms' (2017) 69(1) Administrative LR 83, 105.

[30] Henry McDonald, 'Home Office to scrap "racist algorithm" for UK visa applicants' *The Guardian* (London, 4 August 2020) <https://www.theguardian.com/uk-news/2020/aug/04/home-office-to-scrap-racist-algorithm-for-uk-visa-applicants> accessed 27 December 2020. For further discussion of the risks of discrimination in ADM systems, see further in this Journal, Gianna Seglias, 'Bias and Discrimination in Opaque Automated Individual Risk Assessment Systems: Challenges for Judicial Review Under the Equality Act 2010' (2021) 10 OUULJ 51.

for the digitisation of the civil service from the 1990s. Recently, for instance, several NHS Trusts have allowed Google's DeepMind to access 1.6 million patient records, in response to Google's willingness to develop health analytic tools for NHS use.[31] As with any outsourcing contract, it is essential that accountability mechanisms such as qualitative performance targets for contractors and sanctions for non-compliance are included.[32] This is particularly important considering the well-documented indifference that private sector bodies often have for administrative law concerns including procedural fairness.[33]

2. Regulatory Bodies

In light of the challenges outlined above, regulatory accountability must be obtained prior to deployment. While it will be argued that the optimal method of ADM system regulation is through regulatory tools including data protection laws, impact assessments and codes of practice (Part C), this section outlines

---

[31] Madhumita Murgia, 'NHS trusts sign first deals with Google' (*Financial Times*, 19 September 2019) < https://www.ft.com/content/641e0d84-da21-11e9-8f9b-77216ebe1f17> accessed 27 December 2020.

[32] Regarding possible accountability mechanisms, it is possible to draw analogues with examples given in Pauline Allen, David Hughes, Peter Vincent-Jones, et al., 'Public Contracts as Accountability Mechanisms: Assuring Quality in public health care in England and Wales' (2016) 18(1) Public Management Review 20.

[33] UN Special Rapporteur on extreme poverty and human rights, *Digital technology, social protection and human rights* (Report A/74/493, 2019).

the need for regulatory bodies to achieve regulatory accountability.

A. An Algorithm Commission?

Lord Sales has argued that since government lacks the required technical expertise, an expert 'Algorithm Commission' should be established, which would scrutinise algorithms before deployment, in a manner similar to pre-legislative scrutiny by the Joint Committee on Human Rights.[34] This, he argues, would protect individuals effectively against the risks posed by algorithms.[35] His regulator would operate in a manner similar to the utilities regulators, with a key role in protecting consumers and decision-subjects.[36] He also suggests that it may operate 'as a sort of constitutional court' with the capacity to determine compliance of ADM systems with legal standards.[37]

Such a suggestion, while novel, is not the first of its kind: Andrew Tutt has argued for an American algorithm regulator similar to the Federal Drug Administration, which would balance technological innovation and public safety.[38] This regulator, Tutt suggests, would prevent algorithms from being lawfully deployed until safety testing is completed and disclosure requirements are met. These, he argues, would be adequate safeguards against concerns such as those outlined above. As a form of risk-based regulation, Tutt's model would categorise algorithms based on their complexity, with corresponding levels of scrutiny afforded to them.[39]

---

[34] Sales (n 3) 11-12.
[35] ibid.
[36] ibid 14.
[37] ibid.
[38] Tutt (n 29).
[39] ibid 107.

Certainly, the proposals of Sales and Tutt are important. Significantly, from an institutional perspective, a single, centralised regulator is preferable to establishing myriad individual sectoral regulators, whose competences risk overlap and disjointed approaches to ADM systems; far better to have a unitary, comprehensive approach across government. Centralising expertise also ensures that the public sector retains agency over algorithmic deployment, rather than leaving it all to private sector coders.[40] If this is the case, it would follow that ADM systems are 'more likely to be designed and implemented in lawful, policy-compliant, and accountable ways.'[41] Additionally, establishing such a regulator would likely bolster individual rights, similar to the Information Commissioner's role in upholding individuals' information rights.[42] In this sense, a regulatory body would be helpful in combatting many of the problems identified in the previous section, particularly with regards to legality and contractor accountability.

However, there are three principal concerns related to Lord Sales' particular proposal. First, by suggesting that it could constitute an advisor to the courts, a parliamentary aid, *and* an executive agency, this Commission would cut across the separation of powers. No other body (at least since 2005[43]) has the power to scrutinise government activity and advise on legislative instruments, in addition to enjoying a symbiotic relationship with the courts. Such a power would be tantamount

[40] Freeman Engstrom and others (n 2) 71.
[41] Sales (n 3) 7.
[42] Information Commissioner's Office, Information Commissioner's Annual Report and Financial Statements 2017-18 (HMSO 2018) 11.
[43] The most notable change in the separation of powers was the Lord Chancellor losing their judicial role, see especially Constitutional Reform Act 2005, Part 2 ('Arrangements to modify the office of Lord Chancellor').

to declaring the Commission to be judge, jury and (program) executioner! Given the value-laden nature of ADM decisions, it is doubtful whether an ADM regulator ought to have such wide-ranging powers. It is unlikely that the public would place confidence in an organisation which formally belongs in the executive, yet materially impacts the decisions of the other branches of the state. Such doubts were expressed towards tribunals when they remained part of the Whitehall machine; removing them from the executive altogether was required to allay concerns.[44] Even if we are able to deal with the potential biased or discriminatory application of ADM systems themselves (which has been doubted in the previous section), the appearance of bias in ADM system adjudication is enough to give cause for concern in itself. A more limited regulatory body must be deployed, one which *either* adjudicates *or* administrates.

Second, given the UK's constitutional framework, it is unclear how significant a role the Commission could have in relation to devolution matters: would it have the power, for instance, to reject the implementation of a Scottish ADM system in healthcare, health being a devolved competence? If a devolution matter came before the Supreme Court on the issue, under Lord Sales' approach, we may have the same Algorithm Commission, whose decision is in dispute, advising the Court on the interpretation which ought to be given to the ADM system. This cross-cutting role of the Commission surely cannot be supported. It would be far better to have an ADM system regulator to deal with *narrow* legal questions, within the ADM systems themselves and regarding their application, rather than in the broader constitutional framework in which ADM systems are

[44] Sir Andrew Leggatt, *Report of the Review of Tribunals* (HMSO 2001).

deployed. The institutional expertise of bodies such as the Supreme Court must be respected.

Third, it is unclear whether a regulator could predict how ADM systems will precisely operate at the *ex-ante* stage. By receiving the source code of an algorithm prior to deployment, the Commission could not properly understand the context in which the data would be applied, or which datapoints are more important to the system. Hence, static review of the source code at the *ex-ante* stage is unilluminating;[45] it is far from certain that, just by reading the relevant code, regulators would be able to detect any structural issues, such as the misrepresentation or faulty application of the statutory or policy framework, in an ADM system at this stage.[46] Hence, problems of opacity, as outlined in the previous section, would persist. Far more important would be to identify potential problems before they have the possibility to arise once deployed.

B. A More Limited Regulator

The foregoing analysis is not to suggest that *any* regulator of ADM systems would be so constitutionally problematic. Indeed, a regulatory body ought to be established to ensure the effective deployment of the tools outlined in the next section; every toolkit needs its handyman. It is agreed with Lord Sales that it is sensible to situate the ADM system regulator within the executive, particularly since the executive (rather than, for instance the courts) has the institutional heft which makes it best able to ensure and monitor appropriate policy deployment. Much as the Law Commission is a statutory independent body which is formally part of the executive,[47] an algorithm regulator would be

[45] Kroll and others (n 14) 647.
[46] ibid 651.
[47] See Law Commissions Act 1965.

best suited here, rather than as part of the courts structure. Likewise, it is clear that given the holistic changes driven by AI across government, having several regulators, like the sector-specific 'Ofdogs',[48] would not be particularly effective. The challenges posed by ADM systems are not sector-specific, and do not respect the neat boundaries of government: establishing a single regulatory body, is logical. Where this paper diverges from Lord Sales is regarding the powers and competencies of the regulator. To ensure that the benefits of modern computing may be harnessed by the administrative state, and the risks of such deployment[49] mitigated, a more restricted regulatory agency—provided it is equipped with the right tools—will be the most effective, and least constitutionally troubling, way of regulating ADM systems.

Rather than forming a new government department which, at least for the present, would overstate the importance of ADM systems, a better method would be to follow existing precedent with regards to executive non-departmental public bodies (NDPBs). These provide important examples from which an ADM-specific regulator, charged with deploying the technical toolkit outlined in Part C, may be developed. Two principal case studies can be highlighted. First, the Equalities and Human Rights Commission (EHRC) is empowered under the Equality Act 2006 to promote and enforce equality and non-discrimination laws across Great Britain. With its wide remit, the EHRC has the power to, *inter alia*: provide advice[50] and codes of practice[51] to individuals, organisations and public sector bodies; conduct

[48] Terminology of Carol Harlow and Richard Rawlings, *Law and Administration* (3rd edn, CUP 2009) 67.
[49] As outlined in the first section of this paper.
[50] Equality Act 2006, s 13.
[51] ibid s 14.

investigations into whether unlawful acts have been committed;[52] and institute or intervene in relevant legal proceedings.[53] Many of the Commission's powers, particularly issuing codes of practice and intervening in legal proceedings, can be read across as effective ways to regulate ADM system deployment by government. Particularly in relation to the problem of bias outlined above, an agency modelled on the EHRC could issue guidance to militate against bias and discrimination in the deployment of ADM systems.[54] With powers to issue guidance and advice, the regulator could ensure that legal standards are maintained. At the same time, the reserved power to issue or intervene in legal proceedings as an interested party facilitates *ex-post* accountability: technical tools cannot achieve this on their own. Formally part of the executive, albeit institutionally separate (with for instance a ringfenced budget and autonomy from ministerial control), an ADM system regulator modelled on EHRC lines would be unlikely to cut across the separation of powers: it would be able to intervene in cases concerning ADM systems, advising courts on complex issues when invited, but would not become the ultimate arbiter of ADM system disputes in the symbiotic manner that Lord Sales appears to advocate. This would serve to avoid the appearance of bias which must be avoided in the scrutiny of any government policy, let alone in the scrutiny of inherently opaque algorithmic systems.

The second NDPB from which we can learn is the Information Commissioner's Office (ICO), an NDPB responsible for upholding information rights in the public

---

[52] ibid s 20.

[53] ibid s 30.

[54] The deployment of relevant guidance *ex-ante* may well have ensured that the Home Office ADM system used to filter visa applications (above, text at n 30) may well have avoided having discriminatory effect in the first place.

interest.[55] Its most significant role for present purposes involves supervising adherence to data protection laws;[56] it may investigate data breaches and pursue enforcement measures against breaching organisations. Its powers do not, however, stop with 'hard law' enforcement: the ICO may also audit data controllers, to assess whether they are complying with good practice guidelines.[57] Given the natural intersection between information rights and algorithmic decision-making, the ICO's role can inform substantively the formation of an ADM system regulator. Indeed, if we build the ADM regulatory landscape upon existing technologies like the Data Protection Act (DPA) 2018, using the ICO as a touchstone is sensible, and would serve to further the continuing legality and transparency of ADM systems, within the existing regulatory framework.

Using these examples of executive NDPBs with cross-cutting enforcement and advisory powers, it is likely that a similarly constituted algorithm regulator, established along with the proposed regulatory toolkit, will be effective.

3. A New Regulatory Toolkit

Buttressed by the regulator mooted above, it is through the targeted deployment of legal tools that a more sophisticated regulatory landscape for ADM systems can be engineered.

---

[55] Information Comissioner's Office and Department for Digital, Culture, Media and Sport, *Management Agreement 2018-2021* (ICO 2018).
[56] Data Protection Act 2018, s 115.
[57] ibid s 129.

A. Existing Legislation: GDPR and DPA

Existing legislation provides an effective foundation upon which we can construct the new regulatory landscape. The General Data Protection Regulation (GDPR)[58] governs the use of personal data. Two important provisions are Article 22, which requires safeguards when fully automatic and significant decisions are made about persons, and Article 28, which mandates certain contractual terms between data controllers and processors. As a result, contracted-out ADM systems do not escape scrutiny. The GDPR was implemented in UK law by the DPA 2018, which includes the provision that where decisions are made solely with automated processing, the data controller must 'notify the data subject in writing', who may then 'request the controller to—(i) reconsider the decision, or (ii) take a new decision that is not based solely on automated processing.'[59] Hence, where there are concerns around ADM systems, non-automated alternatives may be found.

By giving individuals alternative means of having decisions made about them, these provisions can be applied to suitably cater for transparency and procedural fairness concerns, thus also militating against the problem of opacity which was outlined earlier: if a decision can be made by multiple pathways guided by the same statutory framework, decision-making processes will become more scrutable. They are, however, limited in scope to *personal* data of *particular* individuals, rather than

[58] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) [2016] OJ L119/1.
[59] DPA 2018, s 14(4).

decision-subjects as a class.[60] A sensible and achieveable suggestion could be to 'bring outsourced ADM which does not involve personal data within its remit' by extending the GDPR framework.[61] By extending the notification and contract terms requirements discussed above, improved data processing and accountability standards would permeate ADM application, further improving the transparency of such systems.

French law serves as a useful template for development here: under the Administrative Code, individuals have a right to an explanation for ADM-processed decisions made about them. In the explanation, they are provided information about the degree to which the decision was automatic, the data processed, and the particular treatment parameters applied to the situation.[62] An approach modelled on the existing data protection framework would ensure that systems are designed with accountability at the centre. This would militate against problems of opacity and legality: the failure of ADM system designers to abide by the GDPR's transparency requirements can be severe.[63]

A further step could be to require user consent for all automated decision-making. However, in order for this to be effective, users would first need to be able to *understand* the system in order to be able to consent to its presence; this is unlikely, given the innate complexity of AI. For instance, in an ADM system governing the allocation of welfare benefits, to be able to

---

[60] Surveillance Studies Network, *Report on the Surveillance Society* (ICO 2006).
[61] Cobbe (n 26) 649.
[62] Digital Republic Act 2016 (France), 'Loi n 2016-1321 pour une République numérique'.
[63] Under the GDPR, the maximum fine for most breaches is the higher figure of either €10 million, or in the case of an 'undertaking', two per cent of its. More severe breaches are liable for twice this: GDPR, Art 83(4)-(5).

understand the hundreds, if not thousands, of relevant variables used by the system to decide to whom to allocate resources; the user would then also need to understand the weight given to particular criteria, and perhaps how their data deviated from the average data of other system users. Given both that 'the digital citizen has little time for data governance',[64] and that attempting to understand mere *simple* algorithms, let alone complex ADM systems with thousands of variables, is incredibly challenging, providing notification to users that they have had a decision made about them through use of an ADM system is more proportionate than requiring consent in every instance. If notified that a decision about them was made electronically, concerned individuals could then request a non-digital alternative decision-making process.[65] Even where an individual does not understand the nature of the decision-making process, providing them with a decision notice and the right to review (as already occurs in welfare benefits applications[66]) is a proportionate step to ensure administrative accountability.

B. Algorithmic Impact Assessments

The European Parliament has recommended that 'algorithms in decision-making systems should not be deployed without a prior

[64] Jonathan A Obar, 'Big Data and The Phantom Public: Walter Lippmann and the Fallacy of Data Privacy Self-management' (2015) 2 Big Data & Society 1, 13.

[65] Suggested by the UN Special Rapporteur on extreme poverty and human rights (n 33) 13.

[66] Claimants have a right to mandatory reconsideration of benefits decisions made about them, see GOV.UK, 'Challenge a benefit decision (mandatory reconsideration)' (2020) < https://www.gov.uk/mandatory-reconsideration> accessed 28 December 2020.

algorithmic impact assessment'.[67] Conducting such an assessment, in a similar manner to the GDPR's data protection impact assessment,[68] would ensure that the most dangerous ADM systems are held back from release, or are subjected to heightened operational scrutiny. Although the operational risk of many ADM systems cannot be comprehensively tested or modelled prior to its application, programmers can raise red flags, which decision-makers can then effectively monitor, so that identified risks do not materialise. In in the event that they do materialise, the impact assessment will have established system-specific procedures to combat their effects. Of course, being able to identify legal risks beyond those inherent in all algorithms requires a degree of legal literacy; this is why it is important for lawmakers, policymakers and technical specialists to learn from one another and understand each other's domains. [69] Directly accounting for the risks outlined in Part A, particularly bias and legality, impact assessments are an essential part of any contemporary regulatory toolkit. As a core tool of *ex-ante* regulation, by militating against risks from emerging in the first place, the *ex-post* accountability burden placed on any regulatory agency would necessarily be less intensive than if there was no such risk-identification process.

C. Codes of Practice

The adoption of codes of practice by public sector bodies would improve the consistency of decision-making and likely improve transparency: such codes are often regarded as a crystallisation of

[67] European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088(INI)) [2020] OJ C449/37 [154].
[68] GDPR, Art 35(7).
[69] Richard Susskind, Tomorrow's Lawyers: An Introduction to Your Future (OUP 2017) 184-190.

good governance norms.[70] The UK Government adopted a Data Ethics Framework in 2018, which is intended to guide public sector 'policy and service design'.[71] Reading this across to ADM systems, it would provide for a proportionate use of data in designing systems, safeguarding that decisions are reached using only relevant considerations. By requiring that these good governance norms are imposed as contractual terms on the contractors who design ADM systems, codes of practice could have normative force and help ensure contractor accountability. This could be further guaranteed by public bodies adopting the recommendations of the House of Commons Science and Technology Committee, that 'Government should produce… a list of where algorithms with significant impacts are being used', and make it available for external analysis.[72] Such a voluntary policy of disclosure is likely to strengthen public confidence in ADM systems, particularly if the list becomes available for scrutiny by third-sector organisations.

One particular duty which could be included in the codes of practice, as proposed by Lord Sales, is a duty to 'have regard' for the interests of decision subjects.[73] Such a wide-ranging duty is desirable, ensuring a wider range of considerations are taken into account by algorithms. Further, it reflects the legislative zeitgeist: the Well-being of Future Generations (Wales) Act 2015 establishes a similar 'take account' duty to all administrative action

---

[70] Cary Coglianese and David Lehr, 'Regulating by Robot: Administrative Decision-Making in the Machine-Learning Era' (2017) 105 Georgetown LR 1147, 1211.
[71] Department for Digital, Culture, Media and Sport, 'Data Ethics Framework' (HMSO 2018) <https://www.gov.uk/government/publications/data-ethics-framework> accessed 20 February 2020.
[72] Algorithms in Decision-Making (HC 351, 2018).
[73] Sales (n 3) 10.

of Welsh public bodies.[74] Looking across to other codes of practice, such as the EHRC guidance on the Equality Act 2010,[75] we see that the use of example scenarios in the guidelines can also be illuminating and ensure decision-makers act according to best practice.

Although codes of practice are non-binding, this is not necessarily a drawback to their efficacy. Compliance with codes of practice is likely to improve public trust in ADM systems and, by incorporating such codes into public contracts for ADM system design, further encourage good faith compliance. It is not the case that hard law enforcement will always be the most effective way of ensuring accountability: when we consider codes of practice alongside existing legislative tools and impact assessments, a hybrid 'carrot and stick' approach is more likely to ensure compliance. Even if it were only government *policy* to ensure that certain good governance norms are required in all public contracts for the design of ADM systems, this would go a long way in improving contractor accountability. Given that the ultimate aim of this toolkit is to improve *ex-ante* accountability, using codes of practice to crystallise norms of good algorithmic governance—both within Whitehall and among its private sector system designers—is likely to result in better-coded ADM systems. This, in turn, will mean that less acute *ex-post* enforcement will be needed.

D. Technical Tools

Fourth, there are particular technical tools available to regulators which can verify automated decisions and ensure compliance with

[74] See especially s 5.
[75] EHRC, 'Equality Act codes of practice' (2019) <https://www.equalityhumanrights.com/en/advice-and-guidance/equality-act-codes-practice> accessed 27 April 2020.

legal standards. Designing ADM systems with external scrutiny in mind must do more than merely examining the inputs and outputs of the system; the ADM system itself must be scrutable. This is particularly important, given the necessary limitations in *ex-ante* scrutiny, and the difficulty in understanding individual decisions unless we understand the ADM system as a whole. Three technical tools in particular are suggested,[76] each of which will serve to promote algorithmic accountability. First, software verification tools allow coders and auditors to test whether an ADM system has certain properties. For instance, in an ADM system designed to implement a statutory scheme, it may be tested using software verification tools whether the system has taken into account all relevant considerations as enumerated in the legislation. Coders will also be able to reason how the system will behave under all conditions:[77] if one of the inputs into the algorithm changes, it would remain possible to verify how it would affect its outputs. Methods of software verification can vary, from exhaustive *ex-ante* testing for all possible variables, to designing a complementary algorithm which will decode ADM systems to prove that the specified properties exist.[78] Testing for particular properties is of particular importance when decision subjects are categorised according to those properties, particularly if these properties are protected characteristics which public bodies must protect.[79] Through this method, we could determine, for instance, whether a given algorithm is likely to give rise to a biased or discriminatory application. In testing any new ADM systems similar to its visa-sorting algorithm,[80] the Home Office could deploy such tests to ensure that it functions in a non-

[76] Drawn from Kroll and others (n 14).
[77] ibid 662.
[78] ibid 663.
[79] Equality Act 2010, Chapters 1 and 2.
[80] See text at n 30.

discriminatory manner. By doing so, Whitehall would continue to make efficiency gains, while having tested comprehensively that the relevant properties are present in its ADM systems.

Second, cryptographic commitments will allow auditors to prove *ex-post* that a given program has performed in accordance with the original underlying commitments, staying true to its legal purpose.[81] These are the digital equivalent of drafting a secure document with a list of particular properties that we wish for a given policy to have; the policy could then be checked later on in relation to this document, to ensure that it has remained true to its undertakings. Whereas parliamentary committees scrutinise the acts of departments, complementary technical tools could be used to test for the legality of ADM systems once they are deployed (and could themselves be used as evidence to parliamentary committees, in their scrutiny of departmental activity). Secure documents would normally be held off-site and checked by an auditor; in the present case this could be facilitated by having the algorithmic regulatory body periodically test ADM systems in relation to these cryptographic commitments. The more detailed the cryptographic commitments that are made in the first place, the more that must be done to ensure fidelity to them as an ADM system operates. This counters the fears that an ADM system may no longer represent the legal framework underpinning the system, thus ensuring its deployment remains lawful, while enabling the source code to remain undisclosed. This further ensures that the software used to implement the policy is determined and recorded prior to the implementation stage, meaning that neither the policy nor its application are affected by external factors.[82] Such tests would remain important whenever an ADM system is deployed in a sensitive area,

[81] Kroll and others (n 14) 666.
[82] ibid 667.

including crime detection and prevention,[83] and the identification of flight risks.[84]

Third, zero-knowledge proofs—a further cryptographic tool—allow the decision-maker to prove that an ADM system has a certain property, without revealing the nature of that property.[85] This is helpful when we consider that transparency *per se* is not necessarily useful. If a decision-maker makes a trio of commitments—to policy, inputs, and a decision reached—then a zero-knowledge proof will allow the verification that the three elements correspond: that the correct policy was applied to the correct input, to reach the stated outcome.[86] This will ensure consistent policy delivery, and facilitate internal departmental auditing, to ensure policies stay congruent with their parent department's remit.

As a result of these technical tools, if applied, ADM systems would become more readily testable and auditable, and their decisions would become easier to explain. All of this could be achieved without compromising the functionality of the systems. By incorporating accountability into system design, it will ensure a more assured *ex-ante* scrutiny of algorithms. At the same time, tools such as these would render unnecessary and disproportionate the broader institutional approach argued for by Lord Sales: by ensuring more effective *ex-ante* regulation of ADM systems as a function of their design, the need to develop a broad public sector regulator with wide ranging adjudicative, advisory and rule-making powers, would be unnecessary.

---

[83] The Durham HART model is a good example of such a model which must remain congruent with its original policy underpinnings. For discussion of the model, see Oswald and others (n 16).
[84] Adapting Veale and Brass (n 4) 138.
[85] Kroll and others (n 14) 668.
[86] ibid.

E. Impact

In addition to supporting *ex-ante* scrutiny of ADM systems, the four regulatory techniques identified would strengthen *ex-post* scrutiny. Internal audit would become more straightforward if cryptographic commitments we used, as these facilitate intra-system testing; administrative accountability would become more automatic; and possible risks could be identified by codes of practice and prophylactically mitigated by impact assessments. All of this is even more assured if an algorithm regulator is deployed. Of course, this regulator, while responsible for ensuring the effective deployment of the toolkit *ex-ante*, would also be able to review *ex-post* any ASM systems developed, considering the prior risk assessment and cryptographic tools. Like the audit capability of the ICO, this would likely strengthen compliance measures. Finally, with powers to institute judicial review proceedings, the regulator, supported with its toolkit, could bring actions, particularly in relation to the risks of bias and discrimination, which are difficult to control for before any contextual data is fed into the ADM system. The crucial distinction between the proposed regulator here, and that of Lord Sales, is twofold. First, the proposed regulator would necessarily be limited, and would necessarily play second fiddle to the primary regulatory *mechanisms* developed. Second, the adjudicative role which was proposed by Lord Sales would not be deployed: it is not for an administrative agency itself to decide on the intrinsic legality of ADM systems; this is primarily a role for the courts. It would be more appropriate for the regulator to issue the (non-binding) industry codes of practice, and ensure scrutiny of ADM systems ex-post; like the EHRC, it may then intervene in legal challenges. This would ensure the regulatory body played an important—but necessarily limited—role in challenges to ADM system adjudication.

Conclusion

In this article, it has been argued that while deploying a regulatory body to ensure accountability for ADM systems may be effective, a superior and more sophisticated method is to apply an expanded regulatory toolkit consisting of existing digital governance rules, algorithmic impact assessments, codes of practice, and cryptographic proof tools. By marrying up existing bodies and procedures with new innovations, accountability techniques may develop to facilitate the beneficial use of ADM systems, rather than inhibiting innovation. Going forward, this framework may serve as a stepping-stone on the path towards comprehensive digital and algorithmic accountability; many of the tools outlined could readily be applied at a local level, and across to the private sector. To avoid the new generation of digital citizens sleepwalking into becoming 'electric sheep',[87] with all major decisions made about our lives concluded by algorithm, regulators must ensure, and citizens must insist on, comprehensive digital accountability. Given the limitless opportunities afforded by AI and MLA, well-crafted regulation offers us the ability to safely harness new technologies, while remaining cognisant of their risks.

[87] Philip K Dick, *Do Androids Dream of Electric Sheep?* (Doubleday 1968).