Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making

*Jennifer Cobbe**

Abstract: The future is likely to see an increase in the public-sector use of automated decision-making systems which employ machine learning techniques. However, there is no clear understanding of how English administrative law will apply to this kind of decision-making. This paper seeks to address this problem by bringing together administrative law, data protection law, and a technical understanding of automated decision-making systems in order to identify some of the questions to ask and factors to consider when reviewing the use of these systems. Due to the relative novelty of automated decision-making in the public sector this kind of study has not yet been undertaken elsewhere. As a result, this paper provides a starting point for judges, lawyers, and legal academics who wish to understand how to legally assess or review automated decision-making systems and identifies areas where further research is required.

Keywords: *public law; administrative law; judicial review; automated decision-making; machine learning; data protection*

* Compliant and Accountable Systems Group, Department of Computer Science and Technology, University of Cambridge (jennifer.cobbe@cl.cam.ac.uk). Many thanks to Jat Singh, Sam Smith, Joe Tomlinson, Swee Leng Harris, Jon Crowcroft, Lauren Downes, Dave Michels, John Morison, Daithí Mac Síthigh, Ross Anderson, and others for advice and for comments on drafts of this paper. Thanks also to the anonymous reviewers. This work was supported by the UK Engineering and Physical Sciences Research Council.

Introduction

The use of automated decision-making ('ADM') systems in the public sector will become increasingly prevalent in future. Decisions involving these systems will need to meet administrative law's standards for public-sector decision-making. However, while work has been undertaken on legal oversight of ADM more generally[1], in other jurisdictions on public sector use of ADM specifically[2], on how Parliament should respond to the growing use of ADM in the UK[3], and on reframing certain principles of English administrative law to highlight risks and challenges in deploying ADM systems[4], it remains unclear how English administrative law will apply to ADM for the purposes of judicially reviewing those decisions. As a result, the courts may be presented with cases involving ADM without a clear understanding of how legal standards for administrative decision-making apply. It's therefore vitally important that work is undertaken to address this deficit. With that in mind, this paper discusses the key and relevant general grounds for judicial review in English administrative law alongside the technical characteristics of ADM systems so as to determine how legal standards can be applied to the use of ADM systems by public bodies[5].

In doing so, this paper does not undertake an in-depth analysis of the finer points of administrative law, of sector-specific statutory requirements, or of the intricacies of ADM systems. Rather, this paper marks a starting point in bridging the gap between the general legal standards for public sector decision-making and the realities of the systems which will

[1] See, e.g., D Keats Citron and F A Pasquale 'The Scored Society: Due Process for Automated Predictions' (2014) 89 *Washington Law Review*; R Binns 'Data protection impact assessments: a meta-regulatory approach' (2017) 7 *International Data Privacy Law* 1; F Doshi-Velez, M Kortz, R Budish, C Bavitz, S Gershman, D O'Brien, S Schieber, J Waldo, D Weinberger, and A Wood 'Accountability of AI Under the Law: The Role of Explanation' (2017) *Harvard Public Law Working Paper No.18-07*. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3064761 [accessed 17/07/2018]

[2] C Coglianese and D Lehr 2017, 'Regulating by Robot: Administrative Decision Making in the Machine-Learning Era' (2017) 105 *Georgetown Law Journal*

[3] A Le Sueur 'Robot Government: Automated Decision-Making and its Implications for Parliament' in A Horne and A Le Sueur (eds) *Parliament: Legislation and Accountability* (Hart Publishing, 2016) p 183

[4] M 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* 2128

[5] Throughout, this paper uses the term 'public body', or 'public bodies', to refer to Ministers, public authorities, local authorities, health authorities, chief constables, reviewable tribunals, regulators, and any other decision-maker which is subject to judicial review when acting in a public law capacity. Note that the Data Protection Act 2018 ('DPA 2018') uses its own definition of 'public body' for the purposes of GDPR (DPA 2018, s 7)

be subject to those standards. In the process, this paper demonstrates that more traditional areas of law can provide a basis for exercising control over the use of new technologies (which are often thought to be specialist in nature or to require entirely new responses).

This high-level approach provides a means for beginning the study of how administrative law should adapt to these forms of decision-making in future. The current law should be understood as a basis for moving forward, rather than as a comprehensive framework which satisfactorily governs public sector ADM. In future, administrative law may need to develop new principles and standards for ADM so as to address some of the issues identified herein, and significant research may be required. As such, as well as applying existing legal standards to ADM, this paper seeks to identify directions for thinking about how administrative law should respond to ADM in a way that makes sense from both a legal and a technical point of view.

The analysis proceeds as follows. First, by discussing ADM itself, including what it is, how it works, and why it poses problems for administrative law and judicial review. Next, by assessing when the use of ADM is permitted; first under data protection law (which applies across the public sector, with some exceptions, and restricts the use of ADM involving personal data), and then common law. Requirements around the information processed in ADM, including relating to relevance and to inferences and predictions produced by ADM systems, are then discussed. Finally, issues of fairness in automated decisions, including non-discrimination and the rule against bias, are considered.

Automated Decision-Making

As this paper intends to apply legal principles to ADM, clarity about what is meant by 'automated decision-making' is important. While ADM does not necessarily include machine learning, this paper primarily refers to decision-making by systems which involve algorithmic processes, including machine learning, to automate human decision-making. In popular discussions these are often termed 'AI', and may also be discussed by reference to

'algorithms' or 'algorithmic decision-making'. There is little publicly-available information on where ADM systems are being or are planned to be used across government, and various public bodies have been reluctant to make this kind of information available[6]. However, research has found that they have been deployed for a number of purposes, including fraud detection, healthcare, child welfare, social services, and policing[7].

Machine learning is the process by which a computer system's statistical model is automatically trained so that it can spot patterns and correlations in (usually large) datasets and infer information and make predictions based on those patterns and correlations[8]. This may involve a practice known as 'profiling'; the processing of data about an individual in order to evaluate personal characteristics relating to their preferences, behaviours, health, economic situation, and so on. ADM systems are generally used in one of two ways. The first involves *solely automated* decision-making; that is, where a system's decision is given effect without human intervention. This contrasts with processes where the system is a guide or one tool among several for a human decision-maker who ultimately brings their judgement to make the final decision themselves.

Machine learning systems are trained using 'training data' (large datasets provided by the system designer). In the supervised machine learning systems commonly used for ADM, the designer also gives the system the desired output of its analysis of that data. In training, the system passes the data through its statistical model to produce a calculated output and then automatically adjusts the internal values (or 'weightings') of that model so as to move the model as a whole incrementally closer to producing the desired output. This process of adjusting weightings is repeated over hundreds, thousands, or millions of iterations until outputs closely match the desired value for the training data.

[6] L Dencik, A Hintz, J Redden, and H Warne 'Data Scores as Governance: Investigating uses of citizen scoring in public services' (2018), p.3. Available at https://datajusticelab.org/data-scores-as-governance [accessed 10/02/2019]

[7] Hintz et al, above n 6

[8] For more in-depth but legally-accessible discussion of how machine learning systems operate, see D Lehr and P Ohm 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51 *U.C. Davis Law Review*; for a deeper dive into machine learning research, see P Domingas 'A few useful things to know about machine learning' (2012) 55 *Communications of the ACM* 10

Once the statistical model has been trained (i.e. its weightings have been determined such that it produces the desired outputs with an acceptable error rate), it can infer information and make predictions based on other data. This involves inputting that data to the system so that it runs through the trained model which ultimately produces the calculated output; an inference or prediction either leading to a decision made by the system itself or upon which a human decision-maker can base their own decision. As this model is constructed by the system designer and then trained on data provided by the designer, the choices made in that process – including in composition of the model, selection of training data, and testing of the system – will have a significant influence on how the system functions and the outputs it produces and thus on the decision-making itself.

Machine learning systems are known to have various issues relating to bias, unfairness, and discrimination in decisions[9], as well as to transparency, explainability, and accountability in terms of oversight[10], and to data protection, privacy, and other human rights issues[11], among others. Much research has sought to improve the standards of ADM systems[12], but this has often not considered legal conceptions or decision-making standards. As a result, the processes and metrics for fair, accountable, and transparent machine learning developed through this research do not always translate easily to legal frameworks. There therefore exist gaps in understanding between technical research and administrative law as well as between the law and the technical characteristics of ADM.

[9] S Barocas and A D Selbst 'Big Data's disparate impact' (2016) 104 *California Law Review*; d boyd and K Crawford 'Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon' (2012) 15 *Information, Communication and Society* 5; V Eubanks *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (Macmillan, 2018)

[10] J Burrell 'How the machine 'thinks': Understanding opacity in machine learning algorithms' (2016) *Big Data & Society*; J A Kroll, J Huey, S Barocas, E W Felten, J R Reidenberg, D G Robinson, and H Yu 'Accountable Algorithms' (2017) 165 *University of Pennsylvania Law Review*; F Pasquale *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015)

[11] R van den Hoven van Genderen 'Privacy and Data Protection in the Age of Pervasive Technologies in AI and Robotics' (2017) 3 *European Data Protection Law* 3; Council of Europe 'Algorithms and Human Rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications' (2017) *Council of Europe study DGI(2017)12*. Available at https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html [accessed 17/07/2018]

[12] Primarily in the 'FAT-ML' – Fairness, Accountability, and Transparency in Machine Learning – research community; see https://www.fatml.org/

Perhaps the greatest challenge relates to the transparency and accountability of machine learning decisions. Explaining decision-making is key to judicial review, but is not always easy with ADM systems in large part because machine learning models typically involve an impenetrable complex of calculations. This problem is often termed 'algorithmic opacity', of which three distinct forms have been identified[13]. The first is *intentional* opacity, where the system's workings are concealed to protect intellectual property. The second is *illiterate* opacity, where a system is only understandable to those who can read and write computer code. And the third is *intrinsic* opacity, where a system's complex decision-making process itself is difficult for any human to understand. More than one of these may combine – for example, a system can be intentionally opaque and it be the case that even if it wasn't then it would still be illiterately or intrinsically opaque. The result of algorithmic opacity is that an automated system's decision-making process may be difficult to understand or impossible to evaluate even for experienced systems designers and engineers, let alone non-technical reviewers. In many cases it will be virtually impossible to determine how or why a particular outcome was reached.

While researchers have sought to address this problem[14], they have not yet succeeded to the extent that solutions – where available – are likely to be useful to a legal or otherwise non-technical audience. Seemingly obvious approaches such as those predicated on revealing the internals of ADM may not produce the expected benefits[15], given that, counter-intuitively, increased transparency over the internal workings of models seems to reduce people's ability to detect even sizeable mistakes[16]. Significant further research is required to determine whether and how best to legally mandate ADM transparency in some

[13] Burrell, above n 10

[14] R Guidotti, A Monreale, F Turini, D Pedreschi, and F Gianotti 'A Survey of Methods For Explaining Black Box Models' (2018) *arXiv preprint*, arXiv:1802.01933. Available at https://arxiv.org/abs/1802.01933 [accessed 17/07/2018]

[15] The benefits of transparency have their limits – see M Ananny and K Crawford 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability' (2016) *new media & society*; L Edwards and M Veale 'Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions?' (2018) 16 *IEEE Security & Privacy 3*

[16] F Poursabzi-Sangdeh, D. G. Goldstein, J. M. Hofman, J. W. Vaughan, and H. Wallach 'Manipulating and Measuring Model Interpretability' (2018) *arXiv preprint*, arXiv:1802.07810. Available at https://arxiv.org/abs/1802.07810 [accessed 11/01/2019]

form, as well as to develop tools for exercising meaningful review[17]. For those lacking a technical understanding of these systems, their decision-making processes may for now remain all but incomprehensible. This poses particular problems for the law. Legal standards and review mechanisms which are primarily concerned with decision-making processes, which examine how decisions were made, cannot easily be applied to opaque, algorithmically-produced decisions. The question therefore arises throughout this paper of how courts and other bodies can assess ADM systems so as to exercise effective review.

Legal responsibility for ADM

While these issues with the complexity and opacity of machine learning are a serious problem, it should be emphasised that ADM systems do not operate autonomously, but under the design and direction of humans. And the law is concerned with the activities of natural or legal persons without directly addressing the actions of machines. Public bodies themselves, rather than machines, therefore remain responsible in law for any decision which involves ADM. This responsibility may take different forms depending on the nature of the unlawfulness in question: for example, a public body may have to account for unlawfully using ADM at all. Or, where using ADM is itself lawful, they may be responsible in law where some feature of a particular ADM system's design or function means that decisions made by or with the assistance of that system are unlawful. The key point is that public bodies are responsible and accountable for the lawfulness of their decision-making whether involving ADM in some way or not, that public bodies are required to meet administrative law's standards when using ADM just as with human decision-making, and that an unlawful decision made by or with the assistance of ADM should be dealt with by reviewers as it would had a similarly unlawful decision been taken by a human[18].

[17] The need for useful tools for those involved in operating or assessing ADM systems has been recognised elsewhere – see M Veale, M Van Kleek, and R Binns, 'Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making' (2018) *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18)*. Available at https://arxiv.org/abs/1802.01029[accessed 17/07/2018]

[18] In another common law jurisdiction, the Australian Government's best practice principles for ADM emphasise that decisions made by or with the assistance of ADM must comply with administrative law (Australian Government 'Automated Assistance in Administrative Decision-Making: Better Practice Guide' (2007), p ix. Available at https://www.oaic.gov.au/images/documents/migrated/migrated/betterpracticeguide.pdf [accessed 13/01/2019])

Given this, in applying administrative law to ADM, what this paper actually discusses is how the law applies to public bodies seeking to use ADM, what kind of considerations arise from their use of ADM, and what questions reviewers should ask to assess decision-making which involves ADM. Even where opacity remains a problem, the law will look to organisational and decision-making processes beyond the algorithm itself. Indeed, despite the relative novelty of ADM systems and their complexity and opacity, many legal questions are more concerned with these non-algorithmic processes. As such, familiar issues which arise in relation to human decision-making are relevant in the same or similar ways in relation to decisions involving machines.

Given that much ADM across the public sector will involve processing personal data, it will at various points be necessary to consider principles, requirements, and restrictions from data protection law – the General Data Protection Regulation ('GDPR')[19] and the Data Protection Act 2018 ('DPA 2018')[20]. In relation to ADM involving personal data[21], public bodies will most likely be acting as a data controller[22] rather than as a data processor[23]. As a result, they will be responsible in law for ensuring compliance with the data protection principles[24], including the obligation to be able to demonstrate compliance with those principles, as well as other data protection requirements[25]. These will be discussed where relevant.

[19] 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, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1 ('GDPR')
[20] As well as providing for clarifications, qualifications, and exemptions from GDPR where permitted, DPA 2018 also extends GDPR to many circumstances where automated-decision making by public bodies is not otherwise covered by GDPR because their activities lie outside the scope of EU law (see DPA 2018, Pt 2 Ch 3; Pt 3; Pt 4)
[21] That is, any information relating to an identified or identifiable natural person (GDPR, art 4(1))
[22] The natural or legal person, public authority, agency or other body which, alone or jointly with others determines the purposes and means of processing (GDPR, art 4(8)). Where the purposes and means of processing are determined by an enactment, the data controller will be the person on whom the obligation to process the data is imposed by that enactment (DPA 2018, s 6(2)) – this will most likely be the public body in question.
[23] GDPR, art 4(8)
[24] GDPR, art 5; see also Recital 39
[25] GDPR, art 5(2)

Review of ADM

There are several noteworthy points in relation to judicial review itself as a process for overseeing ADM. The first relates to how subjects of automated decisions (or their legal representatives) can determine whether a decision which affects them was made unlawfully and so bring judicial review proceedings. Where ADM involves personal data, GDPR may help; an array of information should be provided to those whose personal data is being processed[26], including, in some cases, the existence of ADM and information about the logic involved[27] (the so-called 'right to an explanation'[28]). However, no similar provision exists for ADM not involving personal data.

The three-month time limit normally imposed for issuing judicial review proceedings is also a problem. Due to the complexity of machine learning systems and the quantities of data involved in ADM, this may not be sufficient for a prospective claimant to obtain the data and other information needed to assess a decision, nor may it be sufficient for that assessment to be effectively undertaken. Without reform, the ability of those affected by automated decisions to access justice is at risk. Extending the time limit for judicial review applications in respect of ADM from three to six, nine, or even twelve months would go a significant way towards addressing this problem. Beginning the three-month period from the point when a potential claimant receives the necessary data and information may be an alternative solution.

[26] Processing means "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction" (GDPR, art 4(2))

[27] GDPR, arts 13-14

[28] The existence, extent, and usefulness of this right is much debated. See, e.g., B Goodman and S Flaxman 'European Union regulations on algorithmic decision-making and a 'right to an explanation'' (2016) *2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*. Available at https://arxiv.org/abs/1606.08813 [accessed 17/07/2017]; S Wachter, B Mittelstadt, and L Floridi 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 *International Data Privacy Law* 2; A D Selbst and J Powles 'Meaningful information and the right to explanation' (2017) 7 *International Data Privacy Law* 4; G Malgieri and G Comandé 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 *International Data Privacy Law* 4; L Edwards and M Veale 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 17 *Duke Law & Technology Review*

ADM also differs from human decision-making in that issues which might otherwise be considered appropriate for 'policy' judicial reviews can also be relevant to review of individual decisions (which may be termed 'bureaucratic' judicial review[29]). The fact that individual automated decisions are heavily influenced by the processes and choices around the system (i.e. selection of training data; design and training of models; and testing of systems) means that in order to properly evaluate those individual decisions in a 'bureaucratic' review it may be necessary to also evaluate some of those broader processes and choices[30]. While human decision-makers may be influenced by various legal and non-legal factors, these processes and choices will often be instrumental in determining how systems operate and what outcomes they produce in individual decisions, in a way that is without analogy in humans. These processes and choices can and should be accounted for where this is the case. The distinction between review of policy and review of individual decisions which exists for human decision-making may therefore be significantly blurred or eroded for ADM. Some of the grounds for review discussed herein relate more to review of policies than of individual decisions, and vice-versa, but, in order to exercise effective review of ADM, factors which would otherwise be thought to be outside the scope of a particular challenge may need to be considered.

Finally, it is sometimes thought that computers, generally, and ADM systems, specifically, are inherently rational. This reflects the well-attested psychological phenomenon of *automation bias*, which means that humans are more likely to trust decisions made by machines than by other people and less likely to exercise meaningful review of or identify problems with automated decisions[31]. However, reviewers of ADM should not assume that machines necessarily make better decisions than humans, that machines make decisions which are free from human biases, or that reviewers do not need to exercise the same scrutiny of decisions made by machines as they would of decisions made by humans. ADM systems are engineered by humans, overseen by humans, and used for purposes

[29] See, e.g., P Cane 'Understanding Judical Review and its impact' in M Hertogh and S Halliday (eds) *Judicial Review and Bureaucratic Impact* (Cambridge: Cambridge University Press, 2008); M Elliott and T Thomas 'Tribunal Justice and Proportionate Dispute Resolution' (2012) 71 *Cambridge Law Journal* 2

[30] J Singh, I Walden, J Crowcroft, and J Bacon 'Responsibility & Machine Learning: Part of a Process' (2016). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2860048[accessed 02/08/2018]

[31] L J Skitka, K L Mosier, and M Burdick 'Does automation bias decision-making?' (1999) 51 *International Journal of Human-Computer Studies* 5

determined by humans. Training datasets are constructed by humans, and models are trained to within a particular error rate but not necessarily audited internally or tested across all possible scenarios. As a result, there may be unidentified quirks, flaws, and other problems in a system's model which in certain circumstances result in faulty decisions.

It is therefore quite possible for ADM systems to make decisions which by the law's standards are irrational. The classic statement of irrationality is that it exists where a decision is "so outrageous in its defiance of logic or of accepted moral standards that no sensible person who had applied his mind to the question could have arrived at it"[32]. There is no particular reason why a machine could not fail this test; where a decision would be irrational if it were made by a human, so too will it be irrational where it is made by a machine. Overcoming the assumption that decisions made by machines must be rational, while a psychological step rather than a legal one, is important. Unless reviewers accept that ADM systems can produce irrational results, no assessment of whether an ADM system has in fact produced an irrational result can take place. In reviewing ADM systems, it will therefore be important to hold them to the same standards as humans, lest imperfect systems be permitted to make potentially problematic decisions without the appropriate scrutiny.

Lawfulness of using ADM

In applying legal standards to ADM, the first question to be addressed relates to the circumstances in which it can lawfully be used. Most straightforwardly, decisions will be *ultra vires* in its simplest form when the decision-maker has done something for which they lack legal authority[33]; where this is the case, they will have acted unlawfully whether the decision was taken by automated means or not. Beyond this, there are several further issues to explore in determining whether the law permits a decision to be made by or with the assistance of an ADM system.

[32] *Council of Civil Service Unions v Minister for the Civil Service* [1984] 3 All ER 935; see also *Associated Provincial Picture Houses v Wednesbury Corporation* [1947] 2 All ER 680
[33] See, e.g., *R v Lord Chancellor, ex parte Witham* [1997] 2 All ER 779

The first restrictions on the use of ADM to be considered will be those provided by data protection law, which arise in any situation where personal data is processed in ADM and are therefore general statutory restrictions applicable to many, if not most, areas of public administration[34]. The analysis will subsequently turn to common law questions relevant across the public sector: when using ADM would constitute unlawful sub-delegation by a nominated decision-maker; when using ADM would result in unlawfully fettering discretion; when ADM would be used for improper purposes; when the need to give reasons for a decision precludes the use of ADM; and when the use of contracted-out ADM would be unlawful. Some of these common law principles are supplemented by additional requirements from data protection law where personal data is processed, which will be discussed where relevant.

Use of ADM involving personal data

Under Article 22 GDPR, solely ADM, including profiling, which produces legal or similarly significant effects for the data subject[35] is prohibited unless done on one of three available grounds[36]. Where without a valid legal basis a public body has either made an Article 22 automated decision or has otherwise processed personal data then they have acted unlawfully. Determining whether ADM is caught by Article 22's prohibition will involve answering two questions: whether the decision is 'solely' automated, and whether it would produce legal or 'similarly significant' effects on the data subject.

A decision will clearly be solely automated where the result of ADM is applied directly. But where an automated decision is simply given effect by a human without review or evaluation and without considering other factors then that decision is in fact also solely

[34] Note that DPA 2018 makes specific provision for law enforcement (Pt 3), intelligence services (Pt 4), and other processing which would normally be outside the scope of GDPR (Pt 2 Ch 3)
[35] A natural person who can be identified, directly or indirectly, from personal data (GDPR, art 4(1))
[36] GDPR, art 22; Recital 71; see also Article 29 Data Protection Working Party 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (2018a) 17/EN WP251rev.01, p.19. Available at http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053[accessed 17/07/2018]

automated[37]. To escape Article 22, it is not enough for a human intervener to undertake a cursory or superficial analysis or to simply apply the decision without further consideration. According to the Article 29 Data Protection Working Party[38], "To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the relevant data"[39]. The extent of human intervention should be recorded in the public body's Data Protection Impact Assessment ('DPIA')[40].

The Article 22 prohibition is limited to decisions which produce legal or similarly significant effects concerning the data subject[41]. This has two aspects. The first is relatively straightforward: 'legal' effects arise where the decision in some way affects the data subject's legal rights, including contractual rights[42]. The Working Party has interpreted this to include "cancellation of a contract; entitlement to or denial of a particular social benefit granted by law, such as child or housing benefit; [and] refused admission to a country or denial of citizenship"[43]. The second is 'similarly significant' effects, which could include, for example, the automatic refusal of credit and e-recruitment without human intervention[44]. While not giving objective criteria, the Working Party indicates that decisions akin to those which affect access to health services or education would also likely involve similarly significant effects[45]. Clearly, many decisions made by public bodies are likely to have 'legal or similarly significant effects' concerning the data subject.

[37] Article 29 Data protection Working Party, above n 36, p 20
[38] The Article 29 Data Protection Working Party was an EU advisory body which consisted of representatives of the Data Protection Authorities of each Member State, the European Data Protection Supervisor, and the European Commission. It provided official guidance on the interpretation and application of EU data protection law. It was replaced by the European Data Protection Board (which adopted the work published by the Article 29 Data Protection Working Party) in May 2018.
[39] Article 29 Data Protection Working Party, above n 36, p 21
[40] GDPR, art 35; Recitals 84, 91-94; Article 29 Data Protection Working Party, above n 36, p 21. Data controllers (including public bodies where ADM involves personal data) are required to undertake a Data Protection Impact Assessment in advance of any processing which is likely to pose a high risk to individuals, and particularly that which involves automated processing which produces legal or similarly significant effects (although note that DPA 2018 does not require necessity and proportionality assessments in DPIAs for processing undertaken for law enforcement purposes (s 64))
[41] GDPR, art 22(1).
[42] Article 29 Data Protection Working Party, above n 36, p 21
[43] Article 29 Data Protection Working Party, above n 36, p 21
[44] GDPR Recital 71.
[45] Article 29 Data Protection Working Party, above n 36, p 21

ADM caught by Article 22

Article 22's prohibition is subject to exemptions on three grounds. The first is where the ADM is necessary for the entering into or the performance of a contract between the data subject and the data controller[46]; the second is where the ADM is authorised by law (which must provide suitable safeguards for the data subject's rights, freedoms, and legitimate interests)[47]; and the third is where the ADM is done on the basis of the data subject's explicit consent[48]. If relying on the 'authorised by law' exemption, it is unlikely that a general law authorising a public body to make decisions for a specific purpose but not explicitly authorising ADM and not fulfilling the required conditions would qualify (note that DPA 2018 sets out several obligations for public bodies relying on this exemption[49]). Article 22 ADM is further prohibited by GDPR where it involves a subset of personal data termed 'special category data'[50], with two exemptions[51]. The first exemption involves *explicit* consent under Article 9(2)(a)[52]. The second, for public bodies specifically, is on the basis of Article (9)(2)(g), which applies where processing is undertaken on the basis of law and is necessary for reasons of substantial public interest[53]. The possible bases for Article 22 ADM raise various issues, which will now be discussed.

It is unlikely that public bodies can rely on consent-based exemptions. Consent under GDPR involves a "freely given, specific, informed and unambiguous indication of the data subject's

[46] GDPR, art 22(2)(a); while public bodies are unlikely to enter into contracts with individuals who are using their services, they may do so in the context of employment decisions, for example.
[47] GDPR, art 22(2)(b)
[48] GDPR, art 22(2)(c)
[49] DPA 2018, s 14
[50] 'Special category data' is personal data revealing racial or ethnic origin, political opinions, religious philosophical beliefs, or trade union membership, or the processing of genetic data, biometric data for the purposes of uniquely identifying an individual, data concerning health, or data concerning an individual's sex life or sexual orientation (GDPR, art 9(1))
[51] GDPR, art 22(4)
[52] GDPR, art 9(2)(a)
[53] GDPR, art 9(2)(g); see DPA 2018, s 10, including, in particular, s 10(3) – processing under GDPR, art 9(2)(g) will be lawful only where it meets a condition set out in DPA 2018, Sch 1 Pt 2. Note also that DPA, s 14 places certain requirements on data controllers which rely on art 9(2)(g) in making a solely automated decision which produces legal or similarly significant effects.

wishes"[54]. Whether consent is freely given will depend on whether the provision of a service was conditional upon that consent[55]. However, in most cases, when accessing public services or otherwise submitting to the decision-making of a public body, individuals will have no genuine choice. Indeed, as GDPR puts it:

> "consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller, in particular where the controller is a public authority and it is therefore unlikely that consent was freely given in all the circumstances of that specific situation"[56]

Public bodies should therefore not, as a general rule, make service provision reliant on consent to ADM. Where they do, refusal of consent should not detrimentally affect the individual in question. If consent does not meet GDPR's requirements, then there is no legal basis for processing. The more appropriate legal bases for Article 22 ADM in this context are therefore Articles 22(2)(b) (the decision is authorised by law), and, where processing special category data, 9(2)(g) (processing necessary for reasons of substantial public interest).

Conditions apply to the exemptions allowed for in Articles 22(2)(a) (the decision is necessary for the performance of a contract) and 22(2)(c) (explicit consent), as well as where special category data is being processed. In these cases, there must exist suitable safeguards which protect the rights, freedoms, and legitimate interests of the data subject[57]. In addition, in relation to Article 9(2)(g) (processing necessary for reasons of substantial public interest), the legislation on which this processing is based must itself be proportionate to the aim pursued, respect the essence of the right to data protection, and provide for suitable and

[54] GDPR, art 4(11); see also Recital 32; Article 29 Data Protection Working Party 'Guidelines on consent under Regulation 2016/679' (2018b) 17/EN WP259 rev.01. Available at http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051 [accessed 17/07/2018]; Information Commissioner's Office *Lawful Basis for Processing: Consent* (2018). Available at https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/consent-1-0.pdf [accessed 17/07/2018]

[55] GDPR, art 7(4); Recital 43

[56] GDPR, Recital 43

[57] GDPR, art 22(3)-(4); see also Recital 47

specific measures to safeguard the fundamental rights and interests of the data subject[58]. A general law authorising a public body to make decisions but not explicitly setting out their basis for using ADM would again be unlikely to suffice. If the required safeguards do not exist (whether for ADM involving special category data or otherwise) then the public body lacks a lawful basis for ADM.

If, in undertaking Article 22 ADM, a public body is either processing 'ordinary' personal data under Article 22(2)(a) or is processing special category data under Article 9(2)(g), then determining whether it has legal authority to do so will also involve a necessity test[59]. The key question is whether there exist other effective and less intrusive methods of achieving the same result[60] – i.e. is it *necessary* to employ ADM. Public bodies will need to demonstrate that there are no alternative or more privacy-preserving means of achieving the same outcome[61]. While each decision will stand on its own merits depending on its circumstances, where there are other effective means for making that decision then the necessity test will not be met. If a public body is relying on one of these necessity-based grounds but fails this test then they do not have a lawful basis for ADM.

ADM not caught by Article 22

For ADM which involves personal data but is *not* caught by Article 22, if a public body lacks a legal basis for the processing involved in making that decision then it again lacks the authority to make that decision. This would constitute a failure to comply with GDPR's first data protection principle: that personal data be processed lawfully, fairly, and transparently[62]. Note that data subjects retain a right to object to processing[63], except where this right has been restricted, qualified, or removed by DPA 2018[64]. Where this right

[58] GDPR, art 9(2)(g)
[59] Arising from the fact that these grounds only permit processing where it is necessary.
[60] See Article 29 Data Protection Working Party, above n 36, p 23
[61] Article 29 Data Protection Working Party, above n 36, p 23; see also European Data Protection Supervisor *Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit* (2017). Available at https://edps.europa.eu/sites/edp/files/publication/17-04-11_necessity_toolkit_en_0.pdf [accessed 17/07/2018]
[62] GDPR, art 5(1)(a)
[63] GDPR, art 21
[64] DPA 2018, s 15

exists and has been exercised then the public body lacks a lawful basis for further processing.

There are several grounds on which public bodies may rely for ADM not caught by Article 22, with processing being lawful only if and to the extent that at least one ground applies[65]. The first is the data subject's consent to the processing[66]. Public bodies may also undertake processing where necessary for entering into or the performance of a contract to which the data subject is party[67]. And public bodies may be able to process personal data where doing so is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the body[68]. GDPR also establishes that processing special category data is prohibited unless a specified exemption is met. The available exemptions for public bodies include those which have been discussed already in relation to solely ADM – Article 9(2)(a) (explicit consent) and Article 9(2)(g) (processing necessary for reasons of substantial public interest) – as well as the exemption contained in Article 9(2)(h) for public bodies operating in a healthcare context[69].

If a public body relies on one of the consent bases then the same issues relating to valid consent as discussed previously will arise; it's in many cases unlikely that this will be permitted. If relying on Articles 6(1)(b) (processing necessary for the performance of a contract), 6(1)(e) (processing necessary for the performance of a task carried out in the public interest), 9(2)(g) (processing necessary for reasons of substantial public interest), or 9(2)(h) (processing necessary for various purposes related to healthcare) then the necessity test discussed previously in relation to Article 9(2)(g) will apply. Likewise, if relying on Article 6(1)(e) or Article 9(2)(g) then the same test of the underlying legislation as discussed in relation to Article 9(2)(g) will also apply here. If the public body fails these tests where they apply then they lack a valid legal basis for using ADM.

[65] GDPR, art 6(1); note that public bodies may not rely on the 'legitimate interest' grounds set out in in Article 6(1)(f).
[66] GDPR, art 6(1)(a)
[67] GDPR, art 6(1)(b)
[68] GDPR, art 6(3); see DPA 2018, s 8; this ground can only be relied upon if the processing is undertaken pursuant to EU or domestic law which meets an objective in the public interest and is proportionate to the aim pursued.
[69] GDPR, art 9(2)(h); see also recital 53; DPA 2018, ss 10-11; depending on the circumstances, public bodies may able to process special category data where it is necessary for a variety of healthcare purposes.

Use of ADM by nominated decision-makers

Administrative law establishes that where legislation requires that a decision be made by a particular person (e.g. a Minister), it shouldn't be delegated to others as a means of escaping accountability[70] (although where no particular individual is nominated, decisions may in many cases be taken by other members of the public body[71]). While this rule is primarily concerned with delegating decision-making to another person, it also has implications for ADM.

The key question is whether it is lawful for a nominated decision-maker to make use of an ADM system. Courts have previously held that nominated decision-makers who take *advice* from others have not necessarily delegated their authority to them[72], provided this doesn't amount to the decision-maker having had the decision dictated to them[73] (for example, where they have reserved the right to disagree with the advice[74]). It would therefore likely be the case that a decision-maker cannot rely on an ADM system to effectively make the decision for them, unless this is explicitly provided for in an enactment (indeed, concern over the legality of decisions made by computer led to provision for this being included in the Social Security Act 1998[75]). The Article 29 Data Protection Working Party's 'token gesture' test[76] could be adopted as a guide here. While this was intended for determining whether an automated decision involving personal data is a *solely* automated decision, it also provides a useful test for decisions which do not involve personal data. Adopting this test would establish that the use of ADM would be lawful where a nominated decision-maker can show that they have exercised meaningful oversight of the decision, rather than just a token gesture; that they have the authority and competence to change the decision;

[70] See, e.g., *Noon v Matthews* [2014] EWHC 4330 (Admin); *R v London Borough of Tower Hamlets ex parte Khalique* [1994] 26 HLR 517
[71] *Carltona Ltd v Commissioners of Works* [1943] 2 All ER 560 (CA)
[72] *H Lavender & Son v Minister of Housing and Local Government* [1970] 1 WLR 1231
[73] *Ellis v Dubowski* [1921] 3 KB 621
[74] *Mills v London County Council* [1925] 1 KB 213
[75] Le Sueur, above n 3, pp 188-189; see Social Security Act 1998, s 2
[76] Article 29 Data Protection Working Party, above n 36, p 21

and that they have considered all of the relevant data[77]. Where this test is not met, a nominated decision-maker would have unlawfully delegated their authority to the machine.

However, automation bias is a concern. As previously discussed, people tend to trust decisions made by machines, are more likely to defer to machines, and are less likely to exercise meaningful review of decisions made by machines than if the decision was made by a human. The question of whether a human decision-maker who claims to have relied on an automated system for *advice* has truly exercised meaningful oversight of its decisions will thus be of significant importance. Where an automated decision involves personal data, the public body should have recorded the extent of human intervention in their DPIA. This can help the court assess whether any intervention was truly meaningful. However, this would not provide any assistance for ADM which does not involve personal data. The law may therefore need to develop some means of ensuring that nominated decision-makers can demonstrate that they have not simply given effect to an automated system's decision without the appropriate level of human intervention.

Use of ADM to exercise discretionary powers

Where a decision-maker has a discretionary power, they should take individual circumstances into account when exercising it, they should make each decision on its merits rather than adopting a one-size-fits-all approach, and they should be prepared to depart from policies or guidelines where appropriate. Otherwise they may have acted illegally by fettering their discretion[78] (although public bodies can adhere to policy as a general rule). This will particularly be where decisions involve human rights issues and thus necessarily require discretionary powers to be exercised with due consideration[79]. An immediate concern with ADM is that a decision-maker could fetter their discretion if a particular outcome is recommended to them or they are in some other way guided to make a

[77] This should be reflected in the public body's DPIA if the decision involves personal data or concerns a natural person.
[78] See, e.g., *Padfield v Minister of Agriculture, Fisheries and Food* [1968] 1 All ER 694; *British Oxygen Co Ltd. V Minister for Technology* [1971] AC 610; *R v Warwickshire County Council, ex parte Collymore* [1995] ELR 217; *R (Gujra) v Crown Prosecution Service* [2012] UKSC 52
[79] See, e.g, *R (BBC) v Secretary of State for Justice* [2012] 2012 EWHC (Admin); *R (GC) v Commissioner of Police for the Metropolis* [2011] UKSC 21

particular decision (as was recognised in the Australian Government's best practice principles for the use of ADM systems[80]). Beyond this, the nature of machine learning systems raises further problems.

Typically, machine learning systems uniformly apply a single statistical model to all decisions, in theory producing consistent outputs but not facilitating consideration of the particulars of the case at hand. In some cases this will constitute a prima facie case of fettering discretion. Given this, machine learning systems may be inappropriate for decisions where discretionary powers are likely to need to be exercised on a case-by-case basis, or in other situations where policy may generally be applied but where exceptions are likely to need to be permitted. Since many areas of public administration involve discretionary powers, this is a potentially significant problem for the use of ADM in those areas. It may be the case that their use in such circumstances is unlawful.

However, administrative law is gradually evolving its view on policies, with growing acceptance that consistently applied policy (with appropriate exceptions where necessary to accommodate unusual cases) can provide benefits for good governance, consistency, and predictability[81]. The extent to which ADM systems can help promote these principles through consistently applying policy in circumstances where such an approach is appropriate is therefore a matter for further research (it's worth noting that one stated reason behind providing for decision-making by computer in the Social Security Act 1998 was that it was felt that this could assist in producing consistent decisions[82]). That said, recent developments cast doubt on whether this trend towards preferring consistently applied policy will continue, with equal treatment in the exercise of discretionary powers being cast by the Supreme Court as generally desirable but not amounting to a free-standing principle of administrative law in and of itself[83].

[80] Australian Government, above n 19, p viii, p 37; see also Le Sueur, above n 3, pp 196-197
[81] See, e.g., *R (Lumba) v Secretary of State for the Home Department* [2011] UKSC 12; *Nzolameso v City of Westminster* [2015] UKSC 22
[82] Le Sueur, above n 3, p.198
[83] *R (Gallaher Group Ltd) v The Competition and Markets Authority* [2018] UKSC 25 at [24]-[30]

Use of ADM for improper purposes

The lawfulness of any administrative decision-making will depend on whether powers have been exercised for a purpose for which the public body has legal authority[84]. This applies quite straightforwardly to ADM: a public body will not be permitted to use ADM to make a particular decision where they lack the authority to exercise their decision-making powers for the purpose pursued by that decision. If they lack authority to make decisions for a particular purpose then they lack authority to do so regardless of whether they use ADM in the process or not.

Again, a relevant principle from data protection law further applies this principle to ADM involving personal data. GDPR requires that personal data only be processed for a purpose compatible with that for which it was collected (a principle known as 'purpose limitation')[85]. As with the all of the data protection principles, public bodies as data controllers are responsible for complying with this principle and should be able to demonstrate compliance[86]. As a result, where public bodies otherwise have a valid legal basis to process personal data, they can process that data only for the purpose for which it was collected and for other compatible purposes. Reviewers of ADM may therefore need to determine whether the public body has done so. If this is not the case then the public body has no lawful basis for that processing.

Use of ADM where reasons are required

In administrative law there is no general duty to give reasons for decisions[87]. However, such a duty may be imposed by statute, and the law will usually imply a duty to give reasons in decisions which are judicial or quasi-judicial in nature[88]. For example, reasons may be

[84] See, e.g., *R v Minister for Agriculture, Fisheries and Food, ex parte Padfield* [1968] 1 All ER 694; *R v Secretary of State for Foreign and Commonwealth Affairs, ex parte World Development Movement* [1994] EWHC 1 (Admin); and *Porter v Magill* [2001] UKHL 67
[85] GDPR, art 5(1)(b); see also Recital 50
[86] GDPR, art 5(2)
[87] *R v Secretary of State for the Home Department, ex parte Doody* [1993] 3 WLR 154
[88] *R v Civil Service Appeal Board, ex parte Cunningham* [1991] 4 All ER 310

required in public sector employment decisions[89], in relation to some powers exercised by professional standards and regulatory bodies[90], with the refusal to issue a passport[91], and so on. There may also be a duty to give reasons where the principle of fairness requires it, depending on the circumstances[92]. From this a general rule can be derived that the more serious the decision and its effects, the greater the need to give reasons for it.

In many cases, the use of automated systems will be quite trivial. Whether an automated appointment system operated by a health clinic which deals with minor illnesses or injuries meets the highest standards of decision-making, for example, is, in the grand scheme of things and in most cases, somewhat incidental. But in other scenarios the effects may be rather more profound. ADM systems could potentially be used in many important areas, including policing and criminal justice, healthcare, taxation, welfare provision, social housing allocation, planning, and others. The potential use of these systems spans a whole spectrum of consequence, so the general rule derived from administrative law – that the more serious and consequential a decision the greater the need to give reasons – can be directly applied to ADM.

In doing so, a distinction should be drawn between explanations of *how* a decision was made and reasons for *why* that decision was made. Explanations of how decisions were made would not fulfil an obligation to give reasons[93]. However, just as it is often not straightforward to explain *how* an ADM system reached a particular conclusion, so it is also not straightforward to determine *why* that system reached that conclusion. Where opaque machine learning systems are used to make decisions for which reasons will be required, or even as part of the process of making those decisions, their inexplicability is therefore a serious issue. While there is considerable research into improving the explicability of these systems[94], this is yet to produce useful means for non-technical reviewers to understand

[89] *R v Civil Service Appeal Board, ex parte Cunningham* [1991] 4 All ER 310
[90] *Stefan v General Medical Council* [1999] UKPC 10, [2002] All ER (D) 96
[91] *R v Secretary of State for the Home Department, ex parte Fayed* [1996] EWCA Civ 946, [1998] 1 WLR 763
[92] *R v Higher Education Funding Council, ex parte Institute of Dental Surgery* [1994] 1 All ER 651
[93] See the requirements for reasons set out in *South Buckinghamshire District Council v Porter (No 2)* [2004] 1 WLR 1953 at [36]
[94] R Guidotti, A Monreale, F Turini, D Pedreschi, and F Gianotti, 2018, 'A Survey of Methods For Explaining Black Box Models' (2018) *arXiv preprint*, arXiv:1802.01933. Available at https://arxiv.org/abs/1802.01933[accessed 17/07/2018]

how a decision was made, much less why it was made. As in other situations where machine learning systems are problematic for legal review, further research is required.

The courts might reasonably conclude that the present inability of ADM systems to provide reasons for a decision where necessary should in and of itself be a barrier to the use of these systems for those kinds of decisions in the first place. Some public bodies may attempt to circumvent this barrier by providing retrospective justifications. Courts and other reviewers should be aware of this risk, and should be prepared to exercise the appropriate level of scrutiny when it appears that public bodies are seeking to rely on such justifications[95]. Alternatively, public bodies may attempt to rely on the fact that reasons may not be required where giving them would be particularly difficult or onerous on the decision-maker[96]. The argument could be advanced that the opaque nature of ADM systems makes giving reasons onerous or difficult and thus reasons should not be required. However, this should be resisted as it may result in the use of ADM becoming a means of escaping accountability. At a minimum, where the circumstances require reasons but they cannot be provided, courts should be entitled to conclude that the decision was irrational and therefore unlawful, provided the facts and circumstances indicate that the system should have come to a different result[97].

Use of contracted-out ADM

This concerns situations where a public body contracts[98] with a third-party data processor to undertake ADM, involving personal data[99] or otherwise. Where personal data is involved, GDPR establishes a comprehensive framework governing the relationship between data

[95] See, e.g., *R (Nash) v Chelsea College of Art and Design* [2001] EWHC Admin 538 at [34]; see also *Re Brewster's Application* [2017] UKSC 8 at [50]-[52] (although this was heard on reference from Northern Ireland)
[96] *R v Higher Education Funding Council, ex parte Institute of Dental Surgery* [1994] 1 All ER 651 at [665]-[666]
[97] As they would be entitled to conclude if the decision was made by a human – see, *R v Minister of Agriculture Fisheries and Food, ex parte Padfield* [1968] 1 All ER 694 at [1053]-[1054]; *R v Secretary of State for Trade and Industry and another, ex parte Lonrho plc* [1989] 2 All ER 609 at [620]
[98] For example, as permitted by Deregulation and Contracting Out Act 1994, Pt II or by secondary legislation made under that Act.
[99] For which the public body would act as a data controller.

controllers and data processors[100]. Just as public bodies generally remain responsible and accountable for the quality of contracted-out public services[101], as data controllers they are responsible for compliance under GDPR even where the actual processing is undertaken by a third party[102]. But while issues around the contracts for services delivered by a third party have traditionally been considered to be a private law matter and thus beyond the reach of judicial review[103], GDPR requires that controllers establish certain contractual terms with processors[104]. This potentially provides a means to extend the circumstances in which unlawful sub-delegation occurs to situations where public bodies have not established the required contractual relationship with third-party processors.

While administrative law has so far been reluctant to impose public law standards on private organisations providing contracted-out services[105], extending the remit of review to include contracts between public bodies and third-party data processors does not have that effect. Rather, it imposes a traditional public law requirement on the public body (as a data controller) to meet obligations set out in the applicable legislation (GDPR). Without the required contractual provisions, the public body has not established their relationship with the processor according to the requirements of the legal framework by which that relationship is governed. As a result, the delegation of the decision to the processor (through the delegation of the processing which constitutes the decision) has plainly not occurred lawfully. A court can therefore reasonably find that the public body in question has unlawfully sub-delegated to a third party.

Where a decision doesn't involve personal data, GDPR's framework governing the controller-processor relationship does not apply. The result is that the traditional administrative law position against review of contracts with third parties applies. However,

---

[100] GDPR, arts 24-36; see also Recitals 81-83; Information Commissioner's Office *ICO GDPR guidance: Contracts and liabilities between controllers and processors* (2017). Draft. Available at https://ico.org.uk/media/about-the-ico/consultations/2014789/draft-gdpr-contracts-guidance-v1-for-consultation-september-2017.pdf [accessed 17/07/2018]

[101] R Clayton 'Accountability, Judicial Scrutiny and Contracting Out' (2015) *UK Constitutional Law Blog*. Available at https://ukconstitutionallaw.org/2015/11/30/richard-clayton-qc-accountability-judicial-scrutiny-and-contracting-out [accessed 17/07/2018]

[102] GDPR, art 5(2); art 24

[103] See, e,g., *R v Servite Houses and Wandsworth LBC, ex parte Goldsmith* [2001] LGR 55 (QBD)

[104] GDPR, art 28; Recital 81; this is a new requirement which did not exist in previous legislation.

[105] Clayton, above n 110

as GDPR provides a means to extend review in relation to ADM which does involve personal data, perhaps it is worth considering whether the law should evolve so as to bring outsourced ADM which does not involve personal data within its remit. This may be beneficial where public bodies have not established a legal relationship through a contractual agreement which effectively governs their responsibilities and provides for appropriate oversight mechanisms of a kind comparable to those which exist in a lawful controller-processor relationship[106].

This would continue the trend of recent decades away from respecting the public/private divide and towards an approach to exercising oversight over privately-exercised power which considers the 'nature of the function' being exercised[107]. The alternative seems to be the emergence of two classes of outsourced ADM. The first, involving personal data, would be reviewable where the decision has not been delegated according to GDPR's requirements. The second, not involving personal data, would not be reviewable in the same way. These two classes of decision-making may be equally consequential and may each involve a third party acting on behalf of a public body using the same kinds of systems raising the same kinds of accountability issues discussed throughout this paper. Yet the courts' ability to exercise oversight would wholly differ on the basis of the nature of the data being processed. Such a situation may prove to be untenable given the likelihood of significantly increased public sector use of ADM in future and further research will be needed in order to assess the issues involved and propose a future direction for the law.

Information considered in ADM

Administrative law establishes several requirements around the information considered in decision-making. Decision-makers must not rely on materially-relevant facts which are inaccurate[108]. Further, decision-makers should consider all issues which are relevant to a

[106] Arguments for other approaches in relation to other forms of outsourced public decision-making have also been proposed – see, e.g., C Scott 'Accountability in the Regulatory State' (2000) 27 *Journal of Law and Society* 1

[107] See *R v Panel on Take-overs and Mergers, ex parte Datafin* [1987] 1 All ER 564

[108] See, e.g., *Anisminic Ltd v Foreign Compensation Commission* [1968] 2 WLR 163

decision and should not consider any issues which are not[109]. The data protection principle of 'data minimisation' also gives rise to a further related requirement for ADM involving personal data: that the processed data should be limited to what is *necessary* for the purpose being pursued. These three requirements of accuracy, relevance, and necessity can arise in relation to the data on which the system was trained and to the data inputted to the system in order to produce a decision, as well as to any inferences or predictions produced and considered by the system in the process of making a decision. Where public bodies fail to meet these requirements where applicable, they have made an error either of fact (in relation to accuracy) or of law (in relation to relevance and necessity) which takes them beyond their jurisdiction. These requirements will be explored in more detail.

Training and decision data

For an error of fact to be reviewable it must be materially relevant to the decision in question. This would occur most straightforwardly where the data used in decision-making is inaccurate in some way relevant to the decision. In that case, the public body has made an error of materially-relevant fact and has gone beyond their jurisdiction. Where the decision involves personal data, GDPR's fourth data protection principle ('accuracy')[110] will also be relevant. Public bodies as data controllers are responsible for ensuring the accuracy of personal data and should be able to demonstrate compliance[111].

While human decision-makers may go beyond their jurisdiction by erring in facts materially relevant to a decision, reviewers may need to look beyond this narrow focus with ADM. It may in some cases be necessary to assess the accuracy of the system's training data, which will play a significant role in determining the accuracy of its statistical model and therefore of its inferences and predictions and thus of its decisions. However, while important where inaccuracies in training data may have played a role in a particular decision, this would likely involve reviewing a very large number of records. The practicalities of this may be

[109] See, e.g., *Associated Provincial Picture Houses v Wednesbury Corporation* [1947] 2 All ER 680; *R v Somerset County Council, ex parte Fewings* [1995] 1 WLR 1037; *R (Venables) v Secretary of State for the Home Department* [1998] AC 407
[110] GDPR, art 5(1)(d)
[111] GDPR, art 5(2)

challenging. While technical researchers have proposed ways of easing this to an extent[112], there is not yet one solution which is capable of doing this and which may be of use to those involved in reviewing ADM.

As well as this, in some cases not all of the factors used in training models and making decisions will be directly relevant to a given decision, yet will play a (potentially significant) role in determining its outcome. The relevance of these factors will therefore be an important consideration. There is much overlap with the 'data minimisation' principle for personal data (which holds that personal data should be adequate, relevant, and limited to what is necessary for the purposes for which it is processed[113]). 'Adequate' and 'relevant' map straightforwardly onto the traditional administrative law position that decision-makers should consider all relevant and no irrelevant factors, but 'limited to what is necessary' adds a further requirement. Public bodies would not be permitted to process personal data in ADM unless it is necessary to process that data in order to make the decision; i.e. unless it is impossible to make the decision otherwise.

Problematic here is the use of 'proxies' where systems designers or operators do not wish to use personal details which are particularly sensitive or which relate to characteristics which are protected in some way (for example, relating to gender, ethnicity, sexual orientation, and so on). Machine learning systems may instead be trained on factors which are thought to be a good or reliable proxy for those characteristics. This could mean that decisions are made on the basis of factors which are not themselves directly relevant to or necessary for the decision and without considering factors which are in fact relevant. If this is the case, then the decision may be unlawful.

Two further points should also briefly be mentioned here. The law may require that particular consideration is given to specific factors relevant to a decision. Where an automated system does not do this, because its internal statistical model does not give those factors due weight, it has not applied the law correctly. The law may also require that

[112] See, e.g. C E Brodley and M A Friedl 'Identifying Mislabeled Training Data' (1999) 11 *Journal of Artificial Intelligence Research*
[113] GDPR, art 5(1)(c)

where certain factors are identified a particular outcome should follow. Where the model does not correctly identify these factors or does not proceed to the correct outcome upon doing so, the system will have again erred in law. There are at present no tools which would assist non-technical reviewers here, so research will be required.

Inferences and predictions

Problems also result from the capacity of machine learning systems to infer or predict information from datasets, which may then be considered by the system in producing a decision. The accuracy and relevance of these inferences and predictions will be an important consideration. Even where a system can derive information with 95 percent accuracy, for example, that still means that at least 5 of every 100 decisions will involve inferred or predicted inaccuracies on which the decision may, in part, be based (indeed, a system which is claimed to be 95 per cent accurate may have a false positive rate of over one third[114]). Where inferences constitute personal data, public bodies as data controllers are obliged to ensure that they are accurate[115]; where they do not constitute personal data, the common law position requiring the accuracy of materially relevant facts will apply.

The ability of machine learning systems to infer and predict information can also cause problems in terms of relevance. Just as a reviewer may need to assess whether a system has derived and then considered inaccurate information, it may need to be determined whether it has derived and then considered irrelevant information. If this has occurred then the decision will be unlawful on traditional administrative law principles. Where derivations constitute personal data, GDPR's 'data minimisation' principle further requires that the inferred or predicted information is relevant to the purpose for which the ADM is being undertaken[116]. The same principle also requires that personal data is limited to what is necessary for that purpose. This additional requirement of necessity provides a further limitation on the use of inferences and predictions in ADM, complementing the requirement

[114] D Colquhoun 'An investigation of the false discovery rate and the misinterpretation of p-values' (2014) *Royal Society Open Science*. Available at https://royalsocietypublishing.org/doi/full/10.1098/rsos.140216[accessed 13/01/2019]
[115] GDPR, art 5(1)(d)
[116] GDPR, art 5(1)(c)

of relevance found in both common law and GDPR. Public bodies are thus responsible for ensuring the relevance and (if personal data) necessity of information which is inferred or predicted and then considered in ADM. Where irrelevant or (where applicable) unnecessary information is predicted or inferred and then considered, a finding of illegality should result.

Algorithmic opacity is again a problem for assessing the accuracy, relevance, and necessity of inferences and predictions. There currently exists no means for non-technical reviewers to readily determine whether a system has inferred or predicted and then relied upon inaccurate information. It is also not currently clear how those reviewing ADM could determine whether a system has derived and then relied upon irrelevant information. Requiring public bodies to disclose inferences and predictions made in the process of ADM may be an approach worth considering. However, this would be of limited use in facilitating review of inferences or predictions drawn by a system but not then represented externally in some way. It may be the case that future systems for public sector use should be required to include externalise inferences and predictions in order to facilitate disclosure. Further research here is required.

Fairness in Automated Decisions

Fairness is an active area of research into improving the standards of ADM. Yet while equal treatment and fairness (as a broader principle than procedural fairness) in the exercise of discretionary powers are accepted as being fundamental principles in a democratic society, the Supreme Court has emphasised that they do not translate to justiciable administrative law rights[117]. However, statutory prohibitions on discrimination and the common law rule against bias provide means by which the law seeks, in some circumstances, to promote equality and, to an extent, fairness (broadly conceived of) in decision-making. How these may apply to ADM will be considered in turn.

[117] *R (Gallaher Group Ltd) v The Competition and Markets Authority* [2018] UKSC 25 at [24]-[41]

Non-discrimination

The key principle of the Equality Act 2010 is non-discrimination[118]; both private entities and public bodies are under an obligation to not discriminate on grounds of a protected characteristic[119]. In law, two types of discrimination are recognised. The first is direct discrimination[120], where a decision-maker discriminates against an individual on the basis of a protected characteristic. The second is indirect discrimination[121], where rules which appear to treat everyone equally have the practical effect of excluding or placing onerous requirements on people who share a protected characteristic or disproportionately adversely affecting them when a decision is taken.

Non-discrimination is a fundamental principle of lawful ADM, just as in human decision-making. Relevant technical aspects of ADM should be explored to explain how ADM systems may discriminate. Machine learning systems are trained on large datasets and categorise people as groups of shared characteristics rather than as individuals in order to determine which outcome should be produced. As a result, discrimination between groups is a key aspect of ADM. While much research has focused on issues around bias in training datasets and models as well as fairness of decisions (often expressed in terms akin to actuarial fairness), relatively little work has been undertaken on ensuring that this discrimination is not on grounds of a protected characteristic[122].

The distinction between group-level differences and individual-level behaviour is key. Even if two distinguishable groups of people on the whole behave differently, this does not necessarily say anything about the likely behaviour of any individual member of either group. Indeed, it's often impossible to predict the behaviour of any one individual from knowledge of the collective behaviour of a group to which they belong. Taking a stereotypical example, even if men on the whole tend to watch football more than women

[118] Equality Act 2010, Pt 2 Ch 2
[119] The protected characteristics are age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation (Equality Act 2010, ss 4-12)
[120] Equality Act 2010, s 13
[121] Equality Act 2010, s 19
[122] See, e.g., M Veale and R Binns 'Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data' (2017) *Big Data & Society*

on the whole, knowing this does not tell you anything about how much any individual man or woman watches football. This is a problem for ADM systems, which risk turning group-level differences into discriminatory decisions which affect individuals. And, in law, the problem occurs where a decision itself is discriminatory. The historical practice of car insurance providers charging higher premiums for male drivers provides an analogy. The data on which these decisions were based may have been accurate and women as a whole may have presented lower risk than men as a whole. But, in charging individual men higher premiums than women because of their membership of the group 'men', those companies still unlawfully discriminated on grounds of a protected characteristic[123].

Ultimately, whether an ADM system is discriminatory is a factual question to be answered by reference to the decisions produced by the system in much the same way as for human decision-makers. The nature of the data on which the model was trained, the nature of the model itself, and the nature of the data on which the decision was made, while all potentially relevant to the question of *why* a decision was discriminatory (and potentially relevant to the question of bias, discussed below), are irrelevant in determining *whether* as a matter of law a decision was discriminatory. As such, the issues to be considered in identifying discrimination in automated decision do not materially differ from those which should be considered when identifying discrimination by humans.

The rule against bias

The rule against bias typically applies where a decision-maker has some interest in a case or where they are partial or biased against a subject of a decision in some way. While ADM systems have been proposed as a means for removing bias from decision-making, and while machines themselves do not have an interest in a given decision (as could constitute actual or imputed bias), research has repeatedly shown that these systems can in fact encode biases into decisions[124].

[123] *Association Belge des Consommateurs Test-Achats and Others v Conseil des Ministers* (C-236/09) ECLI:EU:C:2011:100, [2012] 1 WLR 1933

[124] See, e.g., B Friedman and H Nissenbaum 'Bias in Computer Systems' (1996) 14 *ACM Transactions on Information Systems* 3. Available at http://www.nyu.edu/projects/nissenbaum/papers/biasincomputers.pdf [accessed 17/07/2018]; Barocas and Selbst, above n 9; Eubanks, above n 9

Bias may manifest in machine learning systems in a number of ways. For example, where particular groups are or historically were treated less favourably than others by public bodies and this is reflected in the training data, this can produce a model which repeats this difference in treatment. Where particular groups are or were societally disadvantaged and this is reflected in the training data, this can produce a model which repeats the disadvantage. Where the training data was not sufficiently varied for the system to have been trained to adequately handle all possible inputs, this can produce a model which is incapable of dealing with certain inputs equally to others. Or problems may arise where the model simply produces erroneous outputs for certain inputs due to some flaw which was not identified and corrected in testing. As a result, ADM systems may be prone to making decisions which are systematically skewed in some way, rather than acting impartially. This could result in those who meet particular criteria being treated less favourably than those who do not, and may occur in decisions which relate to both natural and legal persons. This could give rise to apparent bias[125].

The courts have previously held that in law bias can arise through "the presence of some factor which could prevent the bringing of an objective judgment to bear, which could distort … judgment"[126]. In ADM, this should include the presence of an internal model which does not produce fair and consistent outputs (for example, a system could, without any intention to do so on the part of the public body, treat those from certain socio-economic backgrounds less favourably than others). That said, while reducing bias is an active area of study in the machine learning research community[127], there is as yet neither consensus on what exactly constitutes bias in ADM nor reliable means for identifying bias or eliminating it from training datasets, models, or automated processes[128] (indeed, some research on reducing bias in machine learning suggests that elimination may be impossible[129]). Nor are

[125] Where a protected characteristic is involved, this could potentially also constitute unlawful discrimination.
[126] *Davidson v Scottish Ministers* [2004] UKHL 34 at [6]; although note that his was a case heard on appeal from Scotland.
[127] See, e.g., R Courtland 'Bias detectives: the researchers striving to make algorithms fair' (2018) 558 *Nature*. Available at https://www.nature.com/articles/d41586-018-05469-3 [accessed 17/07/2018]
[128] Courtland, above n 156
[129] See, e.g., J Kleinberg, S Mullainathan, and M Raghavan 'Inherent Trafe-Offs in the Fair Determination of Risk Scores' (2016) *arXiv preprint*, arXiv:1609.05807. Available at https://arxiv.org/abs/1609.05807 [accessed 30/07/2018]; R Berk, H Heidari, S Jabbari, M Kearns, and A Roth 'Fairness in Criminal Justice Risk Assessments:

there useful tools for non-technical reviewers to reliably determine whether bias exists either in a machine learning system's training data or in its internal statistical model.

However, bias does not need to be proven for apparent bias to arise. The usual test for determining whether apparent bias exists is whether there is 'a real danger of bias'[130], assessed from the viewpoint of a fair-minded and informed observer[131] (although stricter tests may be applied where decision-makers have agreed to be bound by a higher standard[132]). Those reviewing automated decisions may therefore in some case need to determine whether a decision-making system may have encoded a bias into its model which has had an effect on its decisions. If a system produces decisions which consistently benefit or disadvantage a particular group then this possibility is likely to exist.

Conclusions and further research

ADM is likely to be increasingly prominent in the public sector in future. Yet until now there has been little clarity on what the law would require of public bodies in using ADM. This paper has sought to address this deficit by blending various administrative law grounds for judicial review with relevant restrictions and requirements from data protection law and an understanding of the technical features of these systems. In doing so, key questions and issues to be considered by legal reviewers have been identified and discussed. Reviewers should now have some clarity on when a public body has a lawful basis for using ADM. They should know where to begin in assessing the information considered in ADM for accuracy and relevance, both in terms of the training and decision data and of inferences and predictions produced by the system. And they should have an understanding of some things to consider in evaluating ADM for discrimination and bias.

The State of the Art' (2017) *arXiv preprint*, arXiv:1703.09207. Available at https://arxiv.org/abs/1703.09207 [accessed 17/07/2018]; S Corbett-Davies, E Pierson, A Feller, S Goel, and A Huq 'Algorithmic decision making and the cost of fairness' (2017) *arXiv preprint*, arXiv:1701.08230. Available at https://arxiv.org/abs/1701.08230 [accessed 17/07/2018]

[130] *R v Secretary of State for the Environment, ex parte Kirkstall Valley Campaign* [1996]

[131] *Re Medicaments and Related Classes of Goods (No 2)* [2001]; see also *Lawal v Northern Spirit* [2004]

[132] *R v Local Commissioner for Administration in North and North East England, ex parte Liverpool City Council* [1999] All ER (D) 155

Along the way, this paper has highlighted the need for further research in a number of areas, both technical and legal. As noted at several points, two kinds of problem are likely to arise repeatedly in review of ADM. The first of these relates to the fact that transparency remains a general challenge for machine learning systems. The second relates to the more specific challenge of providing means for assessing ADM systems which are useful to non-technical reviewers. While in relation to several of the issues discussed herein there exist academic proposals for technical solutions, these have not yet translated into widely used or easily accessible tools. In order for ADM systems to be used in particularly consequential areas of public administration there will likely need to be some accessible means for providing reasons for decisions. Other developments which would benefit non-technical reviewers of automated systems include means for evaluating the accuracy of training data, means for identifying inferences and predictions to be assessed for accuracy and relevance, and means for assessing bias in machine learning systems.

From a legal point of view, research is needed around the question of sub-delegation, both in terms of when it's appropriate for a nominated decision-maker to delegate to a machine and in terms of the extent to which the courts should exercise oversight where processing which does not involve personal data has been delegated to a third party. There is also scope for work on the extent to which machine learning systems can assist in consistently applying policy where appropriate. More generally, research will be required on the feasibility, benefits, and drawbacks of legally mandating technical transparency or adopting other approaches to permitting more effective review of ADM systems.

In all, while adopting a high-level approach, this paper has established a basis for judges, lawyers, and legal academics to understand how to apply administrative law standards to the public-sector use of ADM systems, while also setting directions for further research.