

Around the Black Box: Applying the Carltona Principle to Challenge Machine Learning Algorithms in Public Sector Decision-Making

Elena Casale*

ABSTRACT

For the first time, important public sector decisions are being taken in the absence of an accountable and identifiable human being. Instead, they are increasingly outsourced to machine learning algorithms (MLAs) to cut costs, save time, and, in theory, improve the quality of decisions made. However, MLAs also pose new risks to fair and legitimate decision making such as bias and rigidity. These risks are often obfuscated by 'intrinsic opacity': the complex interplay between extremely large datasets and code which makes it impossible to trace the decision pathway of an MLA. This 'black box problem' frustrates the review of a public sector decision made by an MLA, as the court is unable to trace the decision-making process and so determine its lawfulness in judicial review. In such cases, it is proposed that the principles of non-devolution surrounding the Carltona principle - the doctrine that allows department officials to exercise powers vested in a minister - offer a promising way of 'getting around' the issue of intrinsic opacity. By conceptualising the outsourcing of a decision to an MLA as an act of devolution, the law can effectively regulate the slippage of democratic accountability that the use of an MLA necessarily entails.

* LL.M (BPP) '21. BA in English (University of Oxford) '20. The author warmly thanks Emma Williams for her guidance and enthusiasm whilst supervising this work.

INTRODUCTION

The daunting consequences of autonomous technology, long apprehended in science fiction, will soon become a routine consideration in judicial review as machine learning algorithms (MLAs) transform the way decisions are made in the UK public sector. The use of MLAs has already been documented in several departments (benefit calculation assessments,¹ child welfare services,² and policing³), and 2020 saw the first challenge to an MLA in court in the *Bridges* case.⁴ At their best, MLAs can be an innovative tool for more efficient public service: cutting costs, raising standards, saving time, and improving the quality of decisions made. However, their use poses new challenges for justice in the form of bias, faulty cross-correlation, inaccuracy, rigidity, automation bias and opacity. When MLAs start to control some of the most important administrative decisions in our lives – ‘which neighbourhoods get policed, which families attain much needed resources, who is short-listed for employment and who is investigated for fraud’⁵ – it is vital that existing structures of judicial review can adequately hold this new form of power to account.

Though even the most advanced MLAs are still far from a state of complete autonomy,⁶ one of their features does allow them to operate beyond human oversight: intrinsic opacity.⁷ Commonly referred to as the ‘black box’

¹ Sarah Marsh, ‘One in Three Councils Using Algorithms to Make Welfare Decisions’ *The Guardian* (London, 15 October 2019) <www.theguardian.com/society/2019/oct/15/councils-using-algorithms-make-welfare-decisions-benefits> accessed 13 October 2021.

² Joanna Redden, Lina Dencik and Harry Warne, ‘Datafied Child Welfare Services: Unpacking Politics, Economics and Power’ (2020) 41 *Policy Studies* 507, 509.

³ Marion Oswald, ‘Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and ‘Experimental’ Proportionality’ (2018) 27 *Information & Communications Technology Law* 223, 223.

⁴ R (*Bridges*) v *Chief Constable of South Wales* [2019] EWHC 2341.

⁵ Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (1st edn, St. Martin’s Press, 2018) 8.

⁶ Tim W Dornis, ‘Of “Authorless Works” and “Inventions without Inventor” – The Muddy Waters of “AI Autonomy” in Intellectual Property Doctrine’ (2021) 43(9) *European Intellectual Property Review* 570.

⁷ Burrell has classified the three types of algorithmic opacity as intentional, illiterate, and intrinsic. Intentional opacity is the obfuscation of whether an algorithm has been used in the first place, and illiterate opacity refers to the fact that most people, including public

problem,⁸ intrinsic opacity is unique to algorithms with a machine learning component, and refers to the complex interplay between extremely large datasets and the code which makes it impossible for even technical experts to trace the decision pathway of an MLA.⁹ This is exemplified by the HART algorithm – Durham Constabulary’s recidivism scoring tool – which incorporates 4.2 million ‘nested and conditionally-dependent decision points.’¹⁰ Faced with just part of the system’s inputs, and the outputs, it is impossible for the decision-maker to decipher the reasoning of the MLA.

This presents a unique reviewability problem in the judicial review of a decision made in full or in part by an MLA: the court is unable to trace the procedure used to make the decision, and so determine its lawfulness. In such cases, it is proposed that the principles of non-devolution surrounding the Carltona principle¹¹ – the doctrine that allows department officials to exercise powers vested in a minister – offer a promising way of ‘getting around’ the issue of intrinsic opacity. By scrutinising whether a decision-maker can lawfully devolve the decision to an MLA, applying the Carltona principle thus avoids the need to understand the ‘black box’ MLA decision.

This article will proceed in three parts. Part I outlines what it is at stake in public sector use of MLAs, framing the risks according to common law principles of unlawful decision making. Part II critically reviews statutory mechanisms to find that they are not sufficient to guard against these risks. Part III then proposes how the Carltona principle could get around the ‘black box’ problem to effectively regulate the ability to use an MLA in public sector decision-making based on three factors: the seriousness of the decision, the

office holders, do not have the requisite technical knowledge to be able to understand the workings of an algorithm even if it is disclosed. Jenna Burrell, ‘How the machine ‘thinks’: Understanding opacity in machine learning algorithms’ (2016) 3 Big Data & Society 1, 1.

⁸ Robert Allen and Dee Masters, ‘In the Matter of Automated Data Processing in Government Decision-Making: Joint Opinion’ (*Cloisters*, 7 September 2019) para 126 <<https://www.cloisters.com/wp-content/uploads/2019/10/Open-opinion-pdf-version-1.pdf>> accessed 11 March 2022; Frank Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information* (Harvard University Press, 2015), 6.

⁹ Jenna Burrell, ‘How the machine “thinks”: Understanding opacity in machine learning algorithms’ (2016) 3 Big Data & Society 1, 5.

¹⁰ Oswald (n 3) 223.

¹¹ *Carltona Ltd v Commissioners of Works* [1943] 2 All ER 560.

technological ability of the MLA, and the level of accountability retained by the decision-maker.

I. THE PROBLEM: MLA RISKS AND CHALLENGES FOR REVIEW

MLAs pose four significant risks to the legitimacy and fairness of the decision-making process: faulty cross-correlation, inaccuracy (comprising bias and discrimination), rigidity, and automation bias. In theory, while these risks can be mapped onto certain established common law illegality grounds, intrinsic opacity precludes such scrutiny and necessitates an alternative framework for scrutinizing MLA decisions.

i) Inaccuracy

The capacity for error is latent in any decision-making process, but the scale of MLA systems means that even small errors give rise to significant cumulative disadvantages across the system.¹² Indeed, the practical reality of MLA inaccuracy has been widely reported for extant MLAs,¹³ and will likely be perpetuated due to the lack of incentives to check for accuracy in the public sector.¹⁴ Furthermore, a certain level of error is inherent to the MLA decision-making process. For example, the MLA used by Durham Constabulary to assess which individuals are at low or moderate risk of recidivism, the Harm

¹² Oscar Gandy, 'Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems' (2010) 12 *Ethics and Information Technology* 29, 5.

¹³ Sonja Starr, 'Evidence-Based Sentencing and the Scientific Rationalization of Discrimination' (2014) 66 *Stanford Law Review* 803, 852; Chai Patel, 'Comments in Lessons from litigation' (Rise of the Robots: Challenging automated decision-making in government Workshop Training, London, 1 July 2021) <<https://learning.publiclawproject.org.uk/lessons/lessons-from-litigation/>> accessed 5 October 2021.

¹⁴ Rebecca A. Williams, 'Rethinking Deference for Algorithmic Decision-Making,' (2018) *Oxford Legal Studies Research Paper* 7/2019, 6 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3242482> accessed 17 October 2021.

Assessment Risk Tool (HART) favours cautious errors, meaning it is more likely to over-predict risk than it is to underpredict.¹⁵

A particularly pernicious form of inaccuracy is MLA bias, which is different from that of the common law rule against bias (the doctrine that a decision-maker must not have a personal interest in a decision).¹⁶ Instead, MLA bias occurs when the system makes predictions or classifications that are systematically too high or low for certain subgroups.¹⁷ This often occurs when the MLA learns bias from inaccurate training data.¹⁸ Amazon's recruiting AI, for example, learned bias against female applicants because past successful hires had been male.¹⁹

While inaccurate training data could in theory be scrutinised in judicial review on the grounds of error of fact, intrinsic opacity means that any errors made by the algorithm as it runs cannot be found easily, if at all.²⁰

ii) Faulty cross-correlation

The MLAs are programmed to find correlation, not causality.²¹ Consequently, characteristics the MLA is not instructed to consider can reappear through proxy; a phenomenon known as faulty cross-correlation. Not only does this mean the MLA is operating autonomously in a sense, outside the bounds of its operational commands, due to the 'black box' problem, it is not possible to work out which correlations have been made post hoc as well. In the

¹⁵ Oswald (n 3) 228.

¹⁶ *R v Bow Street Metropolitan Stipendiary Magistrate, ex parte Pinochet* [1998] UKHL 41.

¹⁷ Partnership on AI, 'Report on Algorithmic Risk Assessment Tools in the US Criminal Justice System' (*Partnership on AI*, 23 April 2019) <www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/> accessed 5 October 2021, 15.

¹⁸ Christopher Knight, 'Automated Decision-Making and Judicial Review' (2020) 25(1) *Judicial Review* 21, 22.

¹⁹ Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women,' *Reuters* (San Francisco, 11 October 2018) <www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-airrecruiting-tool-thatshowed-bias-against-women-idUSKCN1MK08G> accessed 3 October 2021.

²⁰ *E v Secretary of State for the Home Department* [2004] [2004] EWCA Civ 49.

²¹ Brent Mittelstadt and others, 'The ethics of algorithms: Mapping the Debate,' (2016) 1 *Big Data & Society* 1, 5.

recent challenge to the Home Office MLA designed to filter visa applications, for example, it was argued that the MLA discriminated based on race by using nationality as a proxy.²² However, it is important to note that any irrelevant considerations, rather than just protected characteristics such as race or gender, can manifest in consideration in this way. As administrative law requires that a decision-maker does not have to consider irrelevant factors,²³ the MLA could therefore be operating unlawfully.

iii) Rigidity

MLAs apply a single statistical model to decisions,²⁴ which does not necessarily mean a fair outcome for any one person.²⁵ This is because the decision-maker's ability to apply relevant knowledge acquired from experience, interpret complex contextual factors,²⁶ or exhibit ineffable qualities such as mercy,²⁷ is fettered. The decision-maker cannot therefore be sure they are 'keeping their ears open' for relevant factors when applying a policy.²⁸

iv) Automation bias

To assuage these risks, a common strategy for public authorities is to have a 'human in the loop' giving final authorization of the decision to create accountability. However, this practice often just carries forward the risks due to a phenomenon known as 'automation bias': the tendency of non-computer

²² BBC, 'Home Office drops 'racist' algorithm from visa decisions' *BBC* (London, 4 August 2020) <<https://www.bbc.co.uk/news/technology-53650758>> accessed 11 October 2021.

²³ *R v St Pancras Vestry* [1890] 24 QBD 371, 375.

²⁴ Jennifer Cobbe, 'Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making' (2019) 39 *Legal Studies* 1, 647.

²⁵ Jennifer Cobbe, 'Confronting the Algorithmic State,' (*Admin Law Blog*, 24 September 2020) <<https://adminlawblog.org/2020/09/24/jennifer-cobbe-confronting-the-algorithmic-state/>> accessed 14 October 2021.

²⁶ Alexander Babuta and Marion Oswald, *Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges* (RUSI Whitehall Report 2018) 12.

²⁷ Lord Sales 'Algorithms, Artificial Intelligence and the Law,' (2019) 25 *Judicial Review* 46, 52.

²⁸ *British Oxygen v. Minister of Technology* [1971] AC 610; *Associated Provincial Picture Houses v Wednesbury Corporation* [1948] 1 KB 223.

scientists to unquestioningly defer to information supplied by technology.²⁹ Though human oversight may occasionally mitigate the concerns associated with algorithmic decision-making, 'it cannot be systematically relied upon to do so,³⁰ and even where the MLA is only suggestive, it can exert a disproportionate effect on the overall outcome of a decision.³¹ The boundary between a solely automated decision and a partially automated decision is thus blurred, and in practice, the MLA output will be the locus of decision-making. This may be the case for HART, where the Chief Constable of Durham Constabulary has noted that officers using the tool may be reluctant to depart from the automated decision generated.³² A study by Skitka found that automated decisions lowered accuracy rates to 59%, compared with 97% for non-automated decisions, because decision-makers over-relied on faulty automated decisions.³³

As some authors have suggested, the rule against acting under dictation could be applied to prohibit the use of an MLA where it is unquestioningly relied on by the decision-maker. The rule provides that a public body must not mindlessly defer to the opinion of a third party.³⁴ While this could be useful for challenging the use of an MLA where it is thoughtlessly relied on by the decision-maker, it is unlikely to apply where the decision-maker has considered some element of the MLA decision-making personally (e.g. where the decision-maker has been given input on determining the parameters for optimisation or the training data, but is still unable to see the reasoning of the algorithm once it begins machine learning).

As such, while the flaws and risks of MLA decision-making can be mapped conceptually onto established tenets of illegality review, most attempts at

²⁹ Kate Goddard, Abdul Roudsari and Jeremy Wyatt, 'Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators' (2012) 19 *Journal of the American Medical Informatics Association* 121, 121; Knight, (n 18) 21.

³⁰ Karen Yeung, 'Algorithmic Regulation: A Critical Interrogation: Algorithmic Regulation' (2018) 12 *Regulation & Governance* 505, 505.

³¹ Karen Yeung, 'Hypernudge': Big data as a mode of regulation by design' (2017) 20 *Information, Communication & Society* 118, 120.

³² Oswald (n 3) 223.

³³ Linda J. Skitka and others 'Does automation bias decision-making?' (1999) 51(5) *International Journal of Human-Computer Studies* 991, 1002.

³⁴ *H Lavender & Son Ltd v Minister of Housing and Local Government* [1971] 1 WLR 1231.

establishing the illegality of a decision will be frustrated by the intrinsic opacity of MLA reasoning.

II. DEFICITS IN STATUTORY PROTECTIONS

Difficulties in regulating MLAs under the common law prompt recourse to three statutory frameworks: discrimination law under the Equality Act 2010 (EA 2010), information law under the Data Protection Act 2018 (DPA 2018), and privacy law under the Human Rights Act 1998 (HRA 1998). These statutes were invoked in the case of *Bridges*, the first, and so far the only, successful UK court challenge to the use of an MLA. In *Bridges*, the plaintiff, Edward Bridges, sought a successful judicial review against South Wales Police for their use of an automated facial recognition system (AFR Locate). Mr Bridges claimed that the police's use of automated facial recognition technology was unlawful because it breached equality law, rights to privacy, and data protection laws. The grounds of appeal mainly concerned the procedures around MLA operations – such as the Public Sector Equality Duty and Data Protection Impact Assessments – which initially seemed to promise to get around the 'black box' problem. However, an analysis of *Bridges* indicates that these statutes only offer piecemeal protections against the legal risks of MLA decision-making.

The Equality Act 2010

The EA 2010 formed the basis for one ground of appeal in *Bridges*. Mr Bridges argued the police had violated the public sector equality duty (PSED) under s 149(1) of the EA 2010, which requires public authorities to have 'due regard' to the need to eliminate discrimination.³⁵

There were three reasons the Court of Appeal found that the police had failed to comply with the PSED: a 'human in the loop' was not enough to discharge the duty, South Wales Police's statistical analysis of the system was faulty, and the police could not rely on the private manufacturer's view that it was not discriminatory. The PSED is thus a key protection against the risk of discrimination in MLA decisions. But it is only where the relevant group corresponds to individuals possessing a 'protected characteristic'³⁶ that bias

³⁵ Equality Act 2010 (EA 2010), s 149(1).

³⁶ *ibid* ss 4-12.

constitutes discrimination, which is a limited set.³⁷ This was affirmed by Lady Hale in *Coll*,³⁸ where it was held that for discrimination based on a proxy characteristic, there must be ‘exact correspondence’ between the protected characteristic and the ‘disadvantaged class.’³⁹ This significantly limits the scope of the EA 2010 in challenging MLA decisions, which are often biased against classes of people that do not resemble historically protected groups.⁴⁰ For example, where an MLA is biased against lower-income neighbourhoods, it would fall outside the scope of the EA 2010.⁴¹ MLAs will create faulty cross-correlation against random groups that would not be able to avail themselves of the protections of the EA 2010. The EA 2010 is consequently of limited use in challenging algorithmic bias generally, let alone the other legal risks to public-sector MLA use.

The Data Protection Act 2018

Information technology law offers three main protections for MLA decisions. Firstly, section 49 of the DPA 2018 offers protections for decisions that are ‘solely automated’ and have ‘legal or similarly significant effects.’⁴² This provision allows the data subject to request the controller to reconsider the decision or take a new decision not based solely on automated processing.⁴³ Nonetheless, wherever a decision has only been made in part by an MLA, the decision will remain outside the scope of the article.⁴⁴ While there had been some debate which suggested that ‘solely automated’ decisions might encompass decisions made where there has simply not been any ‘meaningful’ human input,⁴⁵ such a phrase did not ultimately appear in the DPA 2018 or the General

³⁷ Allen (n 8) [26].

³⁸ R (*on the application of Coll*) v *Secretary of State for Justice* [2017] UKSC 40

³⁹ *ibid* [28]-[29].

⁴⁰ Sandra Wachter, Brent Mittelstadt and Chris Russell, ‘Why Fairness Cannot be Automated: Bridging the Gap between EU Non-Discrimination Law and AI’ (2021) 41 *Computer Law & Security Review* 1, 6.

⁴¹ *Nederlands Juristen Comité voor de Mensenrechten cs v De Staat der Nederlanden* (case no. C/09/550982/HA ZA 18/388) (District Court of Hague).

⁴² Data Protection Act 2018 (DPA 2018), s 49(1).

⁴³ *ibid* s 49(2).

⁴⁴ Duc Tran, ‘Probing the rules on automated decision making’ (2019) 19 *Privacy & Data Protection* 7, 8.

⁴⁵ HL Deb 13 November 2017, vol 785, col 1870.

Data Protection Regulation (GDPR), and so as it stands, section 49 likely does not protect against decisions made in part by an MLA. This was evident in *Bridges*, where section 49 rights did not form part of the claim. Even if section 49 is triggered and a new decision must be taken, that new decision can still be made in part by an MLA, which largely replicates the risk by way of automation bias.

The second protection offered by the DPA 2018 is the right to the 'knowledge of the reasoning',⁴⁶ which incorporates Articles 13-15, 21 and 22 of the GDPR's right to 'meaningful information about the logic involved.' Whether they offer a full right to an explanation,⁴⁷ or simply that the subject must be told of the existence of the algorithm,⁴⁸ has been the subject of academic discussion. However, in the case of MLAs, this is a moot point: intrinsic opacity means full disclosure would be of limited use.⁴⁹

Finally, the DPA 2018 mandates that data controllers are required to undertake a Data Protection Impact Assessment (DPIA) 'where a type of processing is likely to result in a high risk to the rights and freedoms of individuals.'⁵⁰ In *Bridges*, it was successfully argued that South Wales Police's DPIA was flawed because it failed to consider the rights of persons not on watchlists. While this proved to be successful in *Bridges*, it was held for a quasi-procedural reason (failure to recognise privacy interferences at all), rather than because there was a deficiency in the substantive assessment made for risks that were acknowledged. The level of scrutiny the court would apply where risks have been noted in a DPIA is therefore uncertain,⁵¹ which could be an issue due to practitioners' concerns that public authorities tend to treat DPIA's as a mere

⁴⁶ DPA 2018 (n 42), s 98(1).

⁴⁷ Bryce Goodman and Seth Flaxman, 'EU regulations on Algorithmic Decision-Making and a 'Right to Explanation'' (2017) 38 *AI Magazine* 50, 50.

⁴⁸ Sandra Wachter, Brent Mittelstadt and Luciano 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* 76, 76-83.

⁴⁹ Lilian Edwards and Michael Veale, 'Slave to an Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 *Duke Law & Technology Review* 18, 81.

⁵⁰ DPA 2018 (n 42), s 64(1).

⁵¹ Jack Maxwell and Joe Tomlinson, 'Proving algorithmic discrimination in government decision-making' (2020) 20 *Oxford University Commonwealth Law Journal* 352, 359.

compliance exercise.⁵² Further, it is unclear whether the DPIA's criterion of 'rights and freedoms' encompasses risks that do not fall into established rights categories (such as bias or rigidity), and, in any case, the government has recently proposed to remove the requirement to undertake a DPIA entirely.⁵³

The Human Rights Act 1998

The HRA 1998 Article 8: right to private and family life allows for interferences when they are 'in accordance with the law,' and necessary in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, the protection of health or morals, or the protection of the rights and freedoms of others.⁵⁴ This was a successful ground of appeal in *Bridges*, where the Court of Appeal held that the use of the automated facial recognition MLA was unlawful because there was no clear guidance on where it could be used (the range 'all event types' was 'very broad and without apparent limits').⁵⁵ It is uncertain whether an Article 8 claim would hold where there is clear guidance on who and to what end an MLA can be used. Moreover, the scope of proportionality for the use of MLAs remains to be seen;⁵⁶ in *Bridges*, the proportionality assessment failed. While the appellant argued that the court should consider the privacy interference of all data

⁵² Ravi Naik, 'Comments in Data protection law and automated systems' (Rise of the Robots: Challenging automated decision-making in government Workshop Training, London, 1 July 2021) <<https://learning.publiclawproject.org.uk/lessons/data-protection-law-and-automated-systems/>> accessed 3 November 2021; Megan Goulding, 'Comments in Lessons from litigation' (Rise of the Robots: Challenging automated decision-making in government Workshop Training, London, 1 July 2021) <<https://learning.publiclawproject.org.uk/lessons/lessons-from-litigation/>> accessed 3 November 2021.

⁵³ Department for Digital, Culture Media & Sport, *Data: A new direction* (gov.uk, 10 September

<https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022315/Data_Reform_Consultation_Document__Accessible_.pdf>

59.

⁵⁴ Human Rights Act 1998, art 8.

⁵⁵ *Bridges* (n 4) [130].

⁵⁶ Lorna McGregor, 'Human rights, equality law and automated systems' (Rise of the Robots: Challenging automated decision-making in government Workshop Training, London, 1 July 2021) <<https://learning.publiclawproject.org.uk/lessons/human-rights-equality-law-and-automated-systems/>> accessed 6 November 2021.

subjects cumulatively, the court rejected the argument and held that the proportionality assessment only looks at interferences with each individual's rights.⁵⁷ Therefore, in cases where an MLA collects a limited amount of personal data, but is able to aggregate the data of many people to make an assessment about an individual, it may still fall within the proportionality assessment, despite the scalability of the flaws within the MLA. This might be the case in the application of average data to be determinative of an individual's likelihood to reoffend or risk score, which strips data subjects of agency by confining their futures to predetermined statistics.

III. A POSSIBLE SOLUTION: APPLYING THE CARLTONA PRINCIPLE

We have seen how the 'black box' problem will preclude the ability to scrutinize the intricacies of an MLA's workings to the extent that is necessary for common law judicial review grounds, and while statutory protections offer piecemeal solutions to get around this, they do not comprehensively address the legal challenges stemming from MLAs. Instead, one principle of illegality review could be a promising framework for which to bring a judicial review claim against an MLA: unlawful delegation. As delegation looks at the legitimacy of conferring the decision-making power in the first place, rather than the internal intricacies of the decision-making process, it is a promising way for getting around the 'black box' problem and challenge MLA decision-making holistically.

The law around unlawful delegation protects the presumption that when Parliament vests power in an official or an authority, that power must be exercised by that official or authority rather than discharged to another whom Parliament has not chosen for the task.⁵⁸ The discharge of a decision to another might be seen to map easily onto the use of an MLA, where the synthesising of facts and making of a decision has been outsourced.

In an article on the application of algorithms to American constitutional law, Coglianesi and Lehr acknowledge this 'clear conceptual affinity with the

⁵⁷ Goulding (n 52).

⁵⁸ Matt Hutchings, 'Delegation of Functions: Principles and Recent Perspectives' (2016) 21 *Judicial Review* 93, 93; *Birkdale District Electric Supply Co Ltd v Southport Corp* [1926] AC 355, 364.

spirit and tradition of non-delegation,' but cite three reasons why courts would be unlikely to equate delegation to algorithms with delegation to humans:⁵⁹ (1) machines lack self-interest (they are optimizing for the objectives that those deploying them specify); (2) human governmental officials will retain ultimate control over the specification of algorithms; and (3) algorithms will typically function as legally permissible measurement tools.⁶⁰ While Coglianese and Lehr's critique applies readily to algorithms, these criticisms can be dismissed where the machine learning component exists.

Firstly, faulty cross-correlation and inaccuracy means that MLAs are often not optimising for specified objectives but are rather operating outside the bounds of the parameters initially inputted, changing their inferences based on new data fed into the algorithm. Secondly, human government officials do not retain ultimate control over the mode in which an MLA optimises for a result. MLAs use statistical techniques to learn from data without being explicitly given the instructions on how to do so; continually changing their internal parameters to maximize predictive accuracy.⁶¹ Though there is human involvement at the coding and training stage, once deployed, machine learning can change the way a decision is carried out in a manner imperceptible to the public body.⁶² MLAs cannot express reasoning, which makes them even more removed from the oversight of the decision-maker. Finally, MLAs are fundamentally different from legally permissible measurement tools where delegation has not applied (such as models or calculators⁶³). While a rule-based algorithm is a 'finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given provisions,'⁶⁴ an MLA can, for the reasons already stated, change to exceed the bounds of what it was instructed to do, making and

⁵⁹ Steven Appel and Cary Coglianese, 'Algorithmic Governance and Administrative Law' in Woodrow Barfield (ed), *The Cambridge Handbook for the Law of Algorithms* (Cambridge University Press 2020) 178.

⁶⁰ Cary Coglianese and David Lehr, 'Regulating by Robot: Administrative Decision Making in the Machine-Learning Era' (2017) 105 *Georgetown Law Journal* 1205, 33.

⁶¹ Solon Barocas and Andrew D. Selbst, 'Big Data's Disparate Impact' (2016) 104 *California Law Review* 671, 678.

⁶² Oswald (n 3) 14.

⁶³ Joe Tomlinson and Jack Maxwell, 'Government Models and the Presumption of Disclosure' (Public Law Project, 15 July 2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3652602> accessed 18 October.

⁶⁴ Robin K. Hill, 'What an Algorithm Is' (2015) 29 *Philosophy & Technology* 35, 35.

applying inferences independently and thus substituting in important ways for human judgement.

Delegation or devolution: MLA decision-making and the Carltona principle

The application of delegation principles to MLA decision-making is blocked, however, by the longstanding principle that delegation of a decision involves the transfer of legal accountability for that decision.⁶⁵ We have not yet reached the stage where technology can be responsible at law, which presents difficulties in applying this case law in the contemporary moment. However, there is another line of case law that regulates the transfer of decision-making power where legal responsibility remains with the original decision-maker: otherwise known as devolution.

The paramount case on devolution is *Carltona*,⁶⁶ the case that gave effect to the practical realities of modern government in holding that 'administration of ministerial powers is expected to be carried out by civil servants acting on their behalf and under a level of supervision appropriate for each function.'⁶⁷ Under *Carltona* an act is not delegated, but rather devolved⁶⁸ to officials who act as the 'alter ego'⁶⁹ of the Minister, such that 'a decision made on behalf of a Minister by one of his officials is constitutionally the decision of the Minister himself.'⁷⁰ The *Carltona* principle thus underpins and legitimizes almost the entirety of daily government activity,⁷¹ as a minister - or indeed officials at other levels of the public-sector⁷² - may devolve a power or duty under it.⁷³ It can also be applied

⁶⁵ Mark Campbell, 'The Carltona Doctrine' (2007) 18 Public Law Review 251, 251.

⁶⁶ *Carltona* (n 11).

⁶⁷ *ibid* 563.

⁶⁸ 'Central government: sources of powers' (*Thomas Reuters Practical Law*, 2021) <<http://www.practicallaw.com/2-384-0545>> accessed 11 November 2021.

⁶⁹ Ian Loveland, *Constitutional Law, Administrative Law, and Human Rights: A Critical Introduction* (8th edn, Oxford University Press 2018) 363.

⁷⁰ *Nelms v Roe* [1970] 1 WLR 4 (QB), 8.

⁷¹ Daniel Greenberg, 'Delegation of statutory powers' (*Thomas Reuters Practical Law*, 2021) <<http://www.practicallaw.com/7-383-9733>> accessed 10 November 2021.

⁷² *ibid*.

⁷³ Lord Woolf and others, *De Smith's Judicial Review* (8th ed, Thomas Reuters 2018) ch 5 [186].

to the outsourcing of a decision to an MLA, as the public authority retains legal accountability for a decision that is completely outsourced to an MLA.

The question of whether the Carltona principle could apply to MLA outputs which are then approved by a minister is more complicated. The key for practitioners will lie in establishing that the MLA output was itself a decision that the decision-maker should not have devolved. The MLA output could therefore be challenged as being improperly devolved, even though the minister later approved it. In Australia, what counts as a decision for the purposes of administrative law was elucidated in *Pintarich*⁷⁴ as requiring (1) a mental element, and (2) objective manifestation of the decision.⁷⁵ No such requirements exist in UK law, where the 'decision' reviewable has typically been the final outcome.⁷⁶ However, there is a line of case law that could be instructive on this point. They establish that where a decision-maker's duty or power to act is dependent on a secondary decision, which must also be decided by the decision-maker themselves, the secondary decision cannot be delegated.⁷⁷ This is illustrated by *Gibb*:⁷⁸ where a Minister had a duty to provide accommodation for Romani people, and the Court of Appeal held that this necessarily imported a non-delegable power to determine whether applicants seeking accommodation were, in fact, Romani people.⁷⁹ Similarly, in *Monopolies*,⁸⁰ the House of Lords held that the power to accept a reference depended on an initial decision of whether South Yorkshire was 'a substantial part of the United Kingdom,' which could not be delegated. In *Southwark*,⁸¹ Baroness Hale held the duty to provide a 'child in need' with accommodation entailed a series of judgments which were non-

⁷⁴ *Pintarich v Deputy Commissioner of Taxation* [2018] FCAFC 79.

⁷⁵ Yee-Fui Ng and Maria O'Sullivan, 'Deliberation and Automation – When is a Decision a 'Decision'?' (2019) 26 Australian Journal of Administrative Law 21, 27.

⁷⁶ Lilian Edwards 'Comments on Data protection law and automated systems' (Rise of the Robots: Challenging automated decision-making in government, London, 1 July 2021) <<https://learning.publiclawproject.org.uk/lessons/data-protection-law-and-automated-systems/>> accessed 3 November 2021.

⁷⁷ Azeem Suterwalla and Katherine O'Byrne, 'Discretion and Duty: the Limits of Legality,' *Supperstone, Goudie and Walker on Judicial Review* (6th edn, LexisNexis 2017), 7.13.

⁷⁸ *R v South Hams District Council, ex parte Gibb and Another* (1995) QB 158.

⁷⁹ *ibid* 10.

⁸⁰ *R v Monopolies and Mergers Commission, ex parte South Yorkshire Transport* [1993] 1 WLR 23.

⁸¹ *R (G) v London Borough of Southwark* [2009] 1 WLR 1299.

delegable.⁸² Therefore, the Carltona principle could apply to secondary decisions made by MLAs, where the MLA output is necessarily attached to the primary power.

Ultimately, whether devolution doctrines can apply to MLAs will be a matter for the courts, and one of deep philosophical significance, but their intrinsic opacity and autonomy (which is only set to increase) renders them ripe for consideration on that ground. Crucially, applying the Carltona principle also means applying the limits to devolution established in case law, which offer important protections by regulating when a decision-maker can use an MLA, or when they must make the decision personally.

Limits to devolution: Seriousness

Prior to recent cases, the courts recognised that, absent express statutory language to the contrary, the Carltona principle would generally apply.⁸³ The seriousness of the issues engaged did not matter, even in cases where the exercise of a ministers' powers had a significant impact, such as deciding a life sentence tariff period⁸⁴ or issuing deportation orders.⁸⁵ In *Skinner*,⁸⁶ the Court of Appeal held that the decision to approve a Breathalyzer by the Home Secretary could be devolved, despite the potentially extensive interference the act could have on an individual's liberty if a conviction ensued: 'a vitally important matter might well have occupied the Minister's personal attention... there is in principle no obligation upon the Minister to give it his personal attention.'⁸⁷ The expansive reach of the Carltona principle in these cases at first suggests a worrying permissibility for a minister to devolve decisions to MLAs. However, legislative developments have increased emphasis on the seriousness of the power to limit the implied power to devolve.

In *McCafferty*,⁸⁸ the Court of Appeal set a more open-ended approach to the Carltona principle which marks a departure from *Skinner*. Taking as his

⁸² *ibid* [19].

⁸³ *Re Golden Chemical Products Ltd* [1976] Ch 300, 306.

⁸⁴ *Doody v Secretary of State for the Home Department* [1994] 1 AC 531.

⁸⁵ *R v Secretary of State for the Home Department, ex p Oladehinde* [1991] 1 AC 254, 303.

⁸⁶ *R v Skinner* [1968] 2 QB 700 (CA), 709.

⁸⁷ *ibid* 709.

⁸⁸ *McCafferty's (Terence) Application for Writ of Habeas Corpus* [2009] NICA 59.

starting point that 'it is to be implied that the intention of Parliament is to permit the Carltona principle to apply,' Coughlin LJ set out a number of implied factors to be considered: 'the framework of the legislation, the language of pertinent provisions in the legislation and the importance of the subject matter.'⁸⁹ These were applied in *Adams*,⁹⁰ an appeal by Gerry Adams about whether the order pursuant to which he was interned was valid, given that it had been devolved to the Minister of State rather than made personally by the Secretary of State. It is apparent throughout the judgment that, when determining whether the Carltona principle should apply, the intention of Parliament is paramount;⁹¹ but to determine intention is a 'matter of textual analysis, unencumbered by the application of a presumption,' taking into account the 'gravity of the consequences flowing from the exercise of the power.'⁹² Lord Kerr, giving the unanimous judgment of the court, held that - given the wording of the legislative provisions, and the fact that the Interim Custody Order of the kind used to intern Mr Adams was so 'momentous'⁹³ in its consequences - the intentions of Parliament must have been 'that such a crucial decision should be made by the Secretary of State.'⁹⁴ Applied to MLAs, this would allow the courts to adjudicate based on output on a case by case basis, allowing for MLAs in decisions where the repercussions are menial to allow for administrative efficacy, but limiting them where consequences are serious, such as in predictive policing or recidivism scoring. *Adams* thus offers important protections against the devolution to MLAs based on the nature of the devolved function.

Limits to devolution: Seniority

A further limit to devolution can be found in the case of *Oladebinde*,⁹⁵ the creative application of which could offer another restriction for a public official devolving to an MLA. The Secretary of State for the Home Department had authorized two immigration inspectors to act on his behalf to issue notices of

⁸⁹ *ibid* [17].

⁹⁰ *R v Adams (Northern Ireland)* [2020] UKSC 19, 1 WLR 2077.

⁹¹ John McGarry, 'Carltona - A Matter of Intention' (2020) 25(3) *Judicial Review* 235, 236.

⁹² *Adams* (n 90) 2085.

⁹³ *ibid* 2087.

⁹⁴ *ibid* 2088.

⁹⁵ *Oladebinde* (n 85).

intention to deport the appellants; the issue was whether the Carltona principle applied. Lord Griffiths held that the decision ‘must be taken by a person of suitable seniority in the Home Office for whom the Home Secretary accepts responsibility;’ finding that the inspectors in question were suitably senior given their qualifications and years of experience (22 and 24 years).⁹⁶ The Divisional Court in *DPP*⁹⁷ then expanded this into a more general rule that the Carltona principle requires that the seniority of the official should be appropriate in relation to the nature of the power in question. Though the seniority of an MLA cannot obviously be determined by qualifications or years of experience (though a metric like output consistency could substitute for human experience), these cases could be applied to ascribe importance to ensuring a high level of ministerial credibility for the MLAs used to make decisions, as the minister is ultimately responsible for the output. It is uncertain how far the courts will go to examine the suitability of the algorithm, but at the very least, it would need to satisfy the test of *Wednesbury*⁹⁸ unreasonableness and not be so technologically flawed that no reasonable minister would allow it to exercise the power.⁹⁹

Limits to devolution: Oversight and accountability

The underlying constitutional rationale for the application of the Carltona principle is that ministerial responsibility is retained over the devolved decision, due to the proximate relationship between the Secretary of State and the relevant official.¹⁰⁰ This emphasis on oversight and accountability carries through in the case law, where it is apparent that the Carltona principle will not apply where there is a break in the chain of accountability.¹⁰¹ In *Bourgass*,¹⁰² for example, the court held that a decision by the prison authorities to submit individual inmates to long periods of solitary confinement was unlawful on this basis. The case concerned Prison Rule 45(2) which requires that: 'A prisoner

⁹⁶ *ibid* 300.

⁹⁷ *DPP v Haw* [2007] EWHC 1931.

⁹⁸ *Associated Provincial Picture Houses v Wednesbury Corporation* [1948] 1 KB 223.

⁹⁹ Sedley LJ held that the question of who was suitable was for the official subject to the test of irrationality; *R (on the application of Chief Constable of the West Midlands Police) v Birmingham Magistrate's Court* [2002] EWHC 1087.

¹⁰⁰ Mark Elliot and Robert Thomas, *Public Law* (4th edn, Oxford University Press 2017) 541.

¹⁰¹ McGarry (n 91) 236.

¹⁰² *R (Bourgass) v Secretary of State for Justice* [2015] UKSC 54.

shall not be removed under this rule for a period of more than 72 hours without the authority of the Secretary of State.¹⁰³ The attempt by the Prison Service to claim that the power to take this action had been devolved was dismissed by the Court, on the basis that there was an insufficiently close link between the prison authorities and the ministry to ensure the necessary degree of responsibility and accountability. Lord Reed looked to the overall purpose of the rule – to provide a safeguard for prisoners – and noted that it can only do that if it ensures that prolonged segregation does not continue without being reviewed by independent officials.

As *Bourgass* indicates, or *Sherwin*¹⁰⁴ in the case of agencies, proximity for the purposes of oversight under *Carltona* revolves around the institutional relationships and lines of responsibility between physically distinct structures. Therefore, where a private party has sold an MLA for use in the public sector department, the courts will have to look more closely at the accountability chain between the minister and the MLA. Following Lord Reed's purposive interpretation in *Bourgass*, the court might also look to the overriding purpose of the power to decide whether independent scrutiny by a public authority is necessary. The case law on acting under dictation – which is a form of non-delegation but could also be a separate ground of review – could be useful on this point because it fundamentally addresses the underlying question of whether there is sufficient oversight between a minister and a third party for the purposes of a decision. The classic case on acting under dictation is *Lavender*¹⁰⁵ where it was held that the minister of Housing and Local Government's policy to refuse planning permission if the minister of Agriculture objected amounted to unlawful devolution: Willis J stated that the minister had 'fettered himself in such a way that in this case it was not he who made the decision for which Parliament made him responsible.'¹⁰⁶ The principle derived is that the decision-maker must be shown to be exercising their personal judgment alongside the advice, so they retain control of their overall decision.

¹⁰³ The Prison Rules 1999, s 45(2).

¹⁰⁴ *R v Secretary of State for Social Security ex p Sherwin* [1996] 32 BMLR 1.

¹⁰⁵ *H Lavender & Son Ltd v Minister of Housing and Local Government* [1971] 1 WLR 1231.

¹⁰⁶ *ibid* 1241.

What counts as sufficient independent judgment will inform whether the decision-maker's use of an MLA is lawful. *Audit Commission*¹⁰⁷ is useful on this point when compared with *Lavender*. In *Audit Commission*, the Court of Appeal reasoned that there was no impermissible delegation involved in the Audit Commission adopting the Commission for Social Care Inspectorate's scoring system, as the Audit Commission 'must be taken to have been content with those weightings and to have adopted them.'¹⁰⁸ However, the key reason this case was distinguished from *Lavender*, was that the CSCI's system of weighting was 'transparent and objective',¹⁰⁹ and as a result, it was held the Audit Commission retained sufficient oversight over the rating process. As MLA decisions are intrinsically opaque, it is likely that the court's reasoning would follow *Lavender* to find that reliance on an MLA decision is unlawful. To avoid this, the decision-maker would thus have to effectively make their own decision in parallel until MLAs are able to produce coherent reasoning. This could have been an effective way of the use of a token human failsafe in *Bridges*, or the automation bias seemingly present in the authorities' use of the HART algorithm.

CONCLUSION

As courts grapple with the increasing use of MLAs in public sector decision-making, new ways of thinking about longstanding common law principles of judicial review will undoubtedly emerge. MLAs carry potentially severe risks, and currently, statutory mechanisms for challenging these risks are not comprehensive. The regulatory limits to implied devolution under the Carltona principle could offer significant protections regarding when a decision can be taken by an MLA in the public sector. As machine learning increasingly engenders a slippage in democratic accountability - 'a shift in the command centre of governance, where the democratically accountable officials move to the periphery'¹¹⁰ - the applicability of devolution principles to MLAs is only set to become clearer conceptually. When applied, it could not be assumed that a

¹⁰⁷ *The Audit Commission for England and Wales v Ealing Borough Council* [2005] EWCA Civ 556

¹⁰⁸ *ibid* [26].

¹⁰⁹ *ibid*.

¹¹⁰ Robert Brauneis and Ellen Goodman, 'Algorithmic Transparency for the Smart City' (2017) 20 *Yale Journal of Law and Technology* 103, 117.

statutory authority vested in a senior public servant would also extend to an MLA. Instead, public officials would have to pay closer attention to the serious consequences of an MLA decision, the technological ability of the system, and the structure of oversight and accountability between the official and the MLA. These factors provide important safeguards so that the use of MLAs is possible, but within limits, to ensure that administrative efficiency does not override constitutional propriety. In that sense, the limits to the *Carltona* principle allow the law to ‘go with the flow and channel’ MLAs, rather than merely resisting them.¹¹¹

¹¹¹ Sales (n 27) 47.