

TO WHAT EXTENT IS LEGAL REFORM NEEDED TO OVERCOME THE BARRIERS TO PROVING DISCRIMINATION BY AUTOMATED DECISION MAKING?

*Francis Pascoe*¹

Introduction

Fairness is perhaps the most fundamental value in a democratic society. For a society to flourish, access to opportunities must be abundant and universally available.² Over the past century, the law in England and Wales has developed to expand the opportunities and rights afforded to all members of society. Sadly, there are still a multitude of recent case examples showing that bias and discrimination are affecting individuals today.³ Recent events such as 'Black Lives Matter' and the UK government's 'Levelling Up' agenda have shown how inequalities and discrimination do not just occur as isolated incidents. However, this should not detract from the enormous gains that have been made in equality and decision-making processes.

The exponential growth of computing power and available data over the past few decades is beginning to change the way decisions are made. Automated decision making (ADM) is the use of computer algorithms to make decisions based on data.⁴ It is increasingly being used in many areas of business such as in assessing an individual's credit risk or deciding what content a Netflix user may be interested in.⁵ Advocates for the use of ADM highlight its potential to reduce the waiting times and costs involved in decision making.⁶ Additionally, ADM has the potential to decrease discriminatory decisions. Human decision-makers have

¹ Francis graduated with a 2:1 LLB (Hons) degree in Law

² Centre for Data Ethics and Innovation, 'Review into bias in algorithmic decision-making' (27 November 2020) <<https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making/main-report-cdei-review-into-bias-in-algorithmic-decision-making>> accessed 3 January 2022

³ HM Courts & Tribunals Service, 'Employment tribunal decisions' (5 April 2022) <<https://www.gov.uk/employment-tribunal-decisions>> accessed 3 May 2022

⁴ Centre for Data Ethics and Innovation (n 2)

⁵ Ibid

⁶ Information Commissioners office, 'Rights related to automated decision making including profiling' (01 January 2021) <<https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling>> accessed 2 May 2022

been scientifically demonstrated to carry biases, showing significant variability in their decisions depending on seemingly trivial factors.⁷ One study identified that judges give harsher penalties the more time that elapses from their last meal.⁸ ADM promises to remove this variability from human decision-making. Furthermore, unlike humans, ADM should not carry bias or preconceptions about a data subject. However, poor data collection practices can result in these computers inheriting many of our own biases.⁹

Despite clear benefits, ADM has two main issues. Firstly, it is not trusted by the public. Poling from 2021 found that 38% of Britons were not aware that ADM was being used to make decisions about them personally.¹⁰ Furthermore, 23% of those aware of ADM use disagreed in principle that a 'fair and accurate' algorithm should be used to aid decisions about them.¹¹ This demonstrates a distinct lack of both awareness and trust surrounding ADM. The 2020 exam result news cycle, saw a rise in awareness accompanied by a massive fall in trust surrounding the use of ADM.¹²

The second issue ADM faces is a lack of transparency, often referred to as the 'black box nature of ADM'. This is rightly a cause of the significant lack of trust in ADM. Due to multiple reasons, not least the complexity of algorithms, it is often impossible to determine the 'thought' processes that an algorithm has used to create its output.¹³ This inevitably makes ADM difficult to monitor for faults and as a result, difficult to trust. Additionally, there is an ever-increasing number of examples of ADM making discriminatory decisions due to bad data collection or application.¹⁴ It is impossible to accurately assess the true number of these discriminatory algorithms, but as ADM becomes increasingly available, this issue is highly likely to increase.

While computer scientists are working diligently to create new, innovative ways to reduce the risk of bias and create more transparent algorithms. Many fear that the current opacity in ADM is decreasing the availability of legal remedies for those that have been wronged.¹⁵ It

⁷ Daniel Kahneman et al, *Noise: A Flaw in Human Judgment* (William Collins 2021)

⁸ Shai Danziger & Jonathan Levav, 'Extraneous factors in judicial decisions' [2011] *PNAS* 108(17)

⁹ Laura Carmichael et al, 'Data Mining and Automated Discrimination: A Mixed Legal/Technical Perspective' [2016] *IEEE Intelligent Systems*, 31(6), 51-55

¹⁰ Centre for Data Ethics and Innovation (n 2)

¹¹ *Ibid*

¹² Sean Coughlan, 'Coronavirus: The story of the big U-turn of the summer' (BBC, 13 September 2020) <<https://www.bbc.co.uk/news/education-54103612>> accessed 3 May 2022

¹³ Jennifer Cobbe, 'Administrative law and the machines of government: judicial review of automated public-sector decision-making' (July 2019) *Legal Studies*, 39(4), 640.

¹⁴ Cathy O'Neil, *Weapons of Math Destruction* (New York, Crown Publishers, 2016).

¹⁵ Centre for Data Ethics and Innovation (n 2)

is difficult to detect, let alone prove, that an ADM process has discriminated against an individual. Additionally, there are few legal requirements for data processors to follow regarding testing and monitoring outputs for evidence of bias. Therefore, many organisations may neglect to monitor for discriminatory impacts. Furthermore, there is evidence that the restrictions on ADM put in place by the UK's General Data Protection Regulations (GDPR) may be exacerbating the prevalence of bias by reducing the level of scrutiny applied to ADM.¹⁶

This article aims to explain and evaluate the main legal issues facing algorithmic decision-making processes, in order to determine if the current law is sufficient to protect individuals from discrimination. It will explore the current law in England and Wales surrounding discrimination by ADM, highlighting the key issues that ADM has created in relation to evidence and transparency. It will assess the current legislation on discrimination and how recent case law has started to adapt in relation to ADM. Finally, it will explore the ways that the law could be reformed to better protect vulnerable groups. The article will focus solely on the use of ADM in the public sector for two reasons. Firstly, the public sector has built upon the existing law covering ADM in the private sector. These additional restrictions are far from perfect, but they have undoubtedly reduced the risk of discrimination from ADM. There is little debate that ADM in the private sector poses a real risk of discrimination. Secondly, public sector services are usually monopolistic. Furthermore, the public sector generally has far greater powers to restrict the liberties of individuals. This means that discrimination by the public sector poses a far greater risk to individuals and social cohesion.

1 Current Law on Discrimination and ADM

1.1 A History of Discrimination

Since the beginning of the 20th century, the UK has gradually responded to pressures from equality campaigners and activists, pushing to develop the law to create a fairer society. In 1918, parliament passed the Parliament (Qualification of Women) Act, creating women's suffrage, which was later refined in 1928.¹⁷ In 1965 the UK government then passed a Race Relations Act, banning discrimination on the grounds of race or national origin in public places.¹⁸ This was amended three years later to protect individuals in areas such as housing

¹⁶ Jennifer Cobbe (n 13)

¹⁷ Parliament (Qualification of Women) Act 1918; Representation of the People Act 1928

¹⁸ Race Relations Act 1965

and employment.¹⁹ Although the Race Relations Act had many exemptions that would be deemed unacceptable today, it was the first Act to introduce discrimination protections that resemble those England and Wales have today.

During the next decade, women gained statutory protections. In 1970, the Equal Pay Act created a right to equal contractual pay between the sexes.²⁰ The Sex Discrimination Act 1975 outlawed discrimination against individuals on the bases of their sex.²¹ Finally, in 1995, the Disability Discrimination Act was introduced, bringing in similar protections for those with disabilities to those afforded to women and ethnic minorities.²² These progressive developments have come as the result of great political struggle by many factions in society over the past century. As a result, at the beginning of the 21st century, equality law in England and Wales was comprised of multiple statutes spanning decades.

1.2 Equality Act 2010 and the Public Sector Equality Duty

The Equality Act 2010 was introduced to simplify the law surrounding discrimination by creating one act encompassing all forms of recognised discrimination.²³ The Act also created further protections for some individuals and situations. It extended the protection against indirect discrimination to those with disability as well as increasing the duty to make reasonable adjustments to aid those with disabilities.²⁴ This went some way to correct the controversial precedent set in *Malcolm v Lewisham* just a few years earlier.²⁵ In this case, the House of Lords had ruled that discrimination was not possible where the respondent was not aware of the claimant's disability. The extension of protection against indirect discrimination to those with a disability effectively ended judgments such as these. Amongst other protections, the act also introduced the Public Sector Equality Duty (PSED) which created a positive obligation for public bodies to eliminate discrimination and promote opportunities.²⁶

1.3 Summary of Protected Characteristics

¹⁹ Race Relations Act 1968

²⁰ Equal Pay Act 1970

²¹ Sex Discrimination Act 1975

²² Disability Discrimination Act 1995

²³ Equality Act 2010; Equality and Human Rights Commission, 'Equality Act FAQs' <<https://www.equalityhumanrights.com/en/equality-act/equality-act-faqs>> accessed 23 March 2022

²⁴ *Ibid*

²⁵ *Lewisham LBC v Malcolm* [2008] UKHL 43

²⁶ Equality Act 2010, s 1

Section four of the Equality Act 2010 lists the nine 'protected characteristics' upon which, discrimination is prohibited.²⁷ These are: Age; Disability; Gender reassignment; Marriage and civil partnership; Pregnancy and maternity; Race; Religion or belief; Sex; and Sexual orientation.

The Act outlines three ways in which individuals may be discriminated against based on these characteristics. These are 'direct', 'indirect' and 'combined' discrimination. However, combined discrimination is still prospective and is yet to be made law.²⁸ This has been heavily criticised since the act received royal assent.²⁹

1.4 Discrimination

Direct discrimination occurs when A treats B less favourably due to B having a protected characteristic. There are multiple exceptions and caveats to this, for example, direct discrimination based upon B's age is permissible if it achieves a 'legitimate aim'.³⁰ The aim must be real, objective, and proportionate. It cannot, in itself, be discriminatory and must be significant enough to outweigh the discriminatory effect, for example, in the interest of health and safety. Furthermore, there must be no alternative, non-discriminatory measures available.³¹

Indirect discrimination, defined in section 19, is slightly more complicated. This occurs when "A applies to B a provision, criterion or practice" that is discriminatory.³² A must also apply this same provision to people without the protected characteristic. An example would be a job advert aimed at those within the first five years of their career. As was successfully argued in *Rainbow v Milton Keynes Council*, this would constitute indirect age discrimination.³³ Although this criterion is applied to all applicants; it will likely detriment older people. Without a legitimate aim, this is indirect discrimination. It is one of these two forms of discrimination that claimants must prove when making a discrimination claim against an algorithmic decision-making process. Crucially for discrimination by ADM, this discrimination need not

²⁷ Ibid, s 4

²⁸ Ibid, s 14

²⁹ Catherine Bourne, 'Is it time to legislate for dual discrimination?' (People Management, 2020) <<https://www.peoplemanagement.co.uk/article/1745465/is-it-time-to-legislate-for-dual-discrimination>> accessed 27 June 2022

³⁰ Equality Act 2010, s 13 (2)

³¹ Equality and Human Rights Commission, 'Words and terms used in the Equality Act' (2018) <<https://www.equalityhumanrights.com/en/advice-and-guidance/commonly-used-terms-equal-rights>> accessed 28 July 2022

³² Equality Act 2010, s 19 (1)

³³ *Rainbow v Milton Keynes Council* 1200104/2007

be intentional as was decided in *R (E) v Governing Body of JFS*.³⁴

Combined discrimination, set out in section 14 of the Equality Act, is designed to protect individuals from discrimination due to a combination of protected characteristics.³⁵ For example, a hiring policy that prohibited applications from Muslim women. This would not be discriminating on the grounds of religion as Muslim men could still apply, nor is it discriminating against a sex as an atheistic woman would be able to apply. Section 14 was intended to create a course of redress for individuals affected by this type of discrimination. A 2009 government report estimated that once section 14 was implemented it would increase discrimination cases by an estimated 10%.³⁶

However, in *O'Reilly v BBC 2011*, an employment tribunal considered whether a criterion barring women over the age of 40 from applying for a job would be unlawful.³⁷ The tribunal decided that sex discrimination could be established by comparing a woman to a man over 40. Similarly, age discrimination could be established by comparing a woman to one under the age of 40.³⁸ This means that there is the possibility for an effective claim for combined discrimination without section 14. However, this employment tribunal judgement does not carry the same weight as an appeal decision. Therefore, it remains to be seen whether higher courts will adopt this view, making the need for section 14 mute.

1.5 Public Sector Equality Duty

The Equality Act 2010 also expanded the public sector's positive obligations to promote equality. This concept was first introduced by the 1999 Macpherson Report following the murder of Stephen Lawrence, a black teenager killed in London in 1993.³⁹ Recommendation 70 of the report suggested that the police, and other areas of the government, introduce schemes aimed at "promoting cultural diversity and addressing racism".⁴⁰ A year later the government acted upon this with the Race Relations (amendment) Act 2000.⁴¹ This introduced the Race Equality Duty, the beginning of a move away from rectifying discrimination and towards preventing it. In 2006 a similar positive equality duty was

³⁴ *R (E) v Governing Body of JFS* [2009] UKSC 15, [13]

³⁵ Equalities Act 2010, s 14

³⁶ Editorial, 'The Equality Act 2010 - ten years on' (2020) *IDS Emp Law Brief*, 1130(2), 2

³⁷ *O'Reilly v BBC & Anor* (2011) 2200423

³⁸ Allan Tyrer, 'Multiple discrimination (dual discrimination)' (2012, *Stammering Law*) <<https://www.stammeringlaw.org.uk/disability-equality-law/discrimination/multiple-discrimination>> accessed 27 July 2022

³⁹ Sir William Macpherson, 'The Stephen Lawrence Inquiry' (1999) CM 4262-I

⁴⁰ *Ibid*, 387

⁴¹ Race Relations (amendment) Act 2000

introduced for those with disabilities, followed by a Gender Equality Duty in 2007.⁴²

Section 149 of the Equality Act extended this duty to all protected characteristics and laid out the three aims that all public bodies must strive towards.⁴³ These are to; eliminate discrimination, advance equality of opportunity and to foster good relations with individuals with protected characteristics. The Act expands on these aims, requiring public bodies to remove disadvantages suffered because of their protected characteristics, to meet additional needs of those with protected characteristics as well as to encourage those with protected characteristics to participate in public life, especially where their representation is particularly low.⁴⁴

Finally, section one of the Equality Act creates an obligation for public bodies, when making strategic decisions. They must “have due regard” to reducing “the inequalities of outcome which result from socio-economic disadvantage”.⁴⁵ This socio-economic duty was designed to tackle the huge divide in economic and social opportunity in the UK. However, as with combined discrimination, the government has not enforced it.

1.6 Automated Individual Decision-Making

Since the Equality Act was passed in 2010, there has been an enormous expansion in computing power as well as a significant reduction in costs. This combined with the huge expansion of available data and data processing techniques has allowed humankind to develop more creative and useful algorithms and machine learning models.⁴⁶ Data has become increasingly important to so many organisations and much more available due to innovations such as smartphones, personal voice assistants and social media. This has resulted in increasing concerns over individual privacy.

This prompted the European Union (EU) to introduce the General Data Protection Regulations (GDPR) in 2016.⁴⁷ This replaced the 1995 Data Protection Directive that was introduced when the World Wide Web was just six years old.⁴⁸ The GDPR was a massive update to existing regulations, reflecting the expansion in capabilities in data and computation. In May 2018 GDPR came into force and the UK government implemented this

⁴² Disability Discrimination Act 2005; Sex Discrimination Code of Practice Order 2007

⁴³ Equality Act 2010, s 149

⁴⁴ *Ibid*

⁴⁵ *Ibid*, s 1

⁴⁶ Centre for Data Ethics and Innovation (n 2), 2-4

⁴⁷ General Data Protection Regulation (2016) 2016/679 (EU GDPR)

⁴⁸ Data Protection Directive (1995) 95/46/EC

with the Data Protection Act (DPA) 2018.⁴⁹ The EU regulation gave the government limited scope to make changes. Some of the changes made by the UK were to lower the age of consent to data processing from 16 to 13 and to increase protections for those publishing personal data in the public interest.⁵⁰ The DPA also sought to protect academia by limiting the rights of data subjects to interfere with an organisations ability to process data for scientific, historical, statistical, and archiving purposes.⁵¹

The UK's exit from the EU has complicated the UK's data protection law. Section three of the European Union (Withdrawal) Act 2018 transcribed the EU's GDPR into a new UK GDPR with effect from 1 January 2021.⁵² The new UK GDPR was amended slightly by a 2019 statutory instrument and defined in section 3 and section 205 of the DPA.⁵³ This means that data protection law in the UK must now be read with two legal texts, the Data Protection Act 2018, and the UK GDPR.

1.7 Definition of Personal Data

There are several areas of the UK's data protection law that must be defined before a detailed discussion of the current issues concerning ADM. Direct discrimination, as defined by section 13 of the Equality Act, is impossible without personal data.⁵⁴ The DPA defines "personal data" as "any information relating to an identified or identifiable living individual".⁵⁵ This could include a name or identification number as well as "factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of the individual".⁵⁶ This wide definition of personal data means that all nine protected characteristics are likely to constitute personal data, especially when connected to an individual, something necessary for ADM.

Data relating to certain characteristics has even stronger legal protections. Race, religious belief, and sexual orientation are all classed as 'special category data' and are therefore subject to stricter regulations.⁵⁷ In order to lawfully process special category data, the

⁴⁹ Data Protection Act 2018

⁵⁰ DPO Centre, 'What is the difference between the DPA 2018 and the GDPR? (and why does it matter?)' (2018) <<https://www.dpocentre.com/difference-dpa-2018-and-gdpr/>> accessed 15 July 2022

⁵¹ *Ibid*

⁵² European Union (Withdrawal) Act 2018, s 3

⁵³ The Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019 (S.I. 2019/419), Sch. 2 para. 85(7); Data Protection Act 2018, s 3(10) & s 205(4)

⁵⁴ Data Protection Act 2018, s 13

⁵⁵ *Ibid*, s 3(2)

⁵⁶ *Ibid*, s 3(3)(b)

⁵⁷ UK GDPR, Art 9

processor must identify a legal basis to do so. These are set out in Article 6 of the UK's GDPR.⁵⁸ Additionally, the data processing must comply with one of the conditions found in Article 9 of the GDPR.⁵⁹ Some argue that these restrictions on the processing of special category data have created unintentional problems. Many data processors have chosen not to collect or process such data types to simplify their obligations under GDPR.⁶⁰ Concerns have been raised that this strict regime has decreased the ability of organisations to detect and fight discrimination. As organisations choose not to collect or process such data to comply with the GDPR obligations, they may be missing opportunities to analyse their own practices for discriminatory impact.⁶¹

1.8 Legal or Significant Effects

Section 49 and 50 of the DPA severely restricts the complete reliance on ADM when its decisions produce an adverse legal affect or "significantly affects the data subject".⁶² Solitary reliance on ADM is prohibited unless authorised by law.⁶³ This legal authorisation is often the explicit consent of the individual concerned. However, there are some situations where ADM can be used without the explicit consent of the individual.⁶⁴ Regardless of the legal authorisation, anyone using ADM must ensure that certain safeguards are in place.⁶⁵ They must ensure that the individual is informed about the processing, that the individual can challenge the decision and finally, the processor must conduct regular checks to ensure the system is working correctly.⁶⁶

The phrase 'legal or a similarly significant affect' is relatively ambiguous.⁶⁷ The UK's Information Commissioners Office provides some guidance, based on that of the EU's Article

⁵⁸ *Ibid*, Art 6

⁵⁹ *Ibid*, Art 9

⁶⁰ Frederik Borgesius, 'Strengthening legal protection against discrimination by algorithms and artificial intelligence' (2020) *International Journal of Human Rights*, 24(10), 1572-1593

⁶¹ *Ibid*

⁶² Data Protection Act 2018, s 49(2)(b)

⁶³ *Ibid*, s 49(1)

⁶⁴ Information Commissioners office, 'When is consent appropriate?' (2022) <<https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/when-is-consent-appropriate>> accessed 11 July 2022

⁶⁵ Data Protection Act 2018, s 50

⁶⁶ Information Commissioner's Office, 'Rights related to automated decision making including profiling' (2022) <<https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/>> accessed 4 July 2022

⁶⁷ *Ibid*, s 49 & s 50

29 Working Party, as to what qualifies as a 'significant decision'.⁶⁸ The Working Party identified the cancellation or change to a contract as an example of a legal effect. Other examples included the denial of social benefits or citizenship.⁶⁹ A 'similarly significant affect' is more complicated. The working party identified several guideline criteria on what decisions may produce such an affect. These decisions were ones that could significantly change the behaviour or choices of the data subject; that would have a permanent or long impact; or those that could lead to the exclusion of an individual.⁷⁰ The working party also gave examples of such decisions. Those that affect financial circumstances, access to healthcare, employment opportunities and education.⁷¹

Article 4 of the UK GDPR places similar restrictions on profiling, a type of ADM. The GDPR defines profiling as "automated processing of personal data" used to "evaluate certain personal aspects relating to a natural person".⁷² Profiling is often used to predict many factors, such as an individual's performance at work or financial situation. Profiling has previously been controversially used; one such example is a United States credit firm using profiling to identify those most likely to default on their loans by looking at the shops they used their cards in.⁷³ This can have the effect of tarnishing the credit scores of all individuals from a particular neighbourhood. Despite its strong predictive powers, profiling is a particularly controversial use of ADM as it groups people together by characteristics. This can limit individual autonomy and perpetuate societal inequalities.

2 Opacity in ADM

Algorithmic decision making (ADM) encompasses many different types of decision-making practices. In the modern age, this is always done using a computer. The simplest form of ADM is a coded script instructing a computer that "if X, then Y". However, for ADM to predict or decide that which humans struggle, the algorithms must become much more complex. These algorithms will take in many pieces of information and will search through it to find

⁶⁸ European Commission, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (2018) <<https://ec.europa.eu/newsroom/article29/items/612053/en>> accessed 20 July 2022

⁶⁹ *Ibid*, 21

⁷⁰ *Ibid*, 21-22

⁷¹ *Ibid*, 22

⁷² UK GDPR, Art 4(4)

⁷³ Karen Hao, 'The coming war on the hidden algorithms that trap people in poverty' (2020, MIT Technology Review) <<https://www.technologyreview.com/2020/12/04/1013068/algorithms-create-a-poverty-trap-lawyers-fight-back/>> accessed 2 August 2022

patterns that a human cannot. These algorithms, with the ability to surpass human intelligence, are commonly referred to as Artificial Intelligence (AI). The more complex of these AIs have the ability to learn through trial and error over millions of iterations. This is known as machine or deep learning.

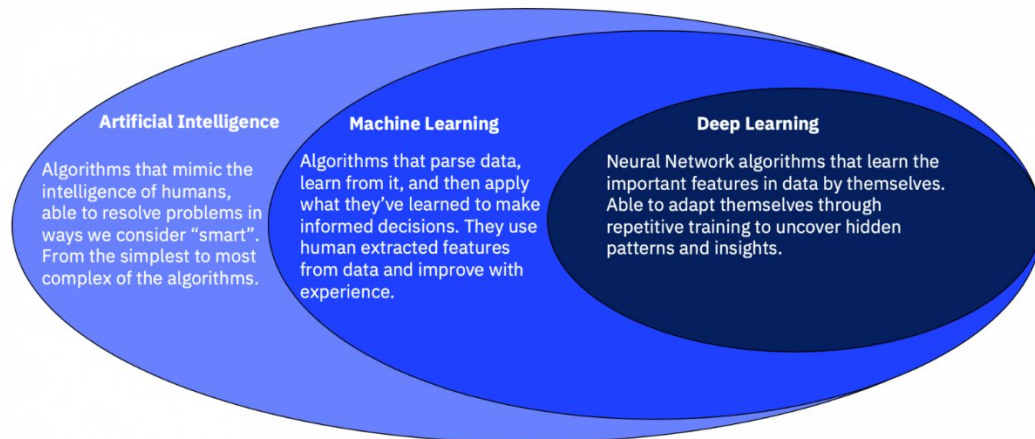


Figure 1 - AI types⁷⁴

Unlike humans, these algorithms tend to specialise in one task and as a result, they are unable to communicate their reasoning behind a decision.⁷⁵ This can cause issues when attempting to prove that an algorithm has relied upon a protected characteristic in its decision-making.⁷⁶ This forms part of a broader issue of algorithmic opacity. Cobbe has identified three main reasons for which algorithmic opacity arises; intentional, illiterate, and intrinsic opacity.⁷⁷ Intentional opacity occurs when an algorithms' code, data or reasoning is protected or withheld due to commercial interests such as the protection of intellectual property. This can lead to a lack of scrutinization and transparency resulting in discriminatory practices not being recognised. Commercial interests can also create issues when an algorithm has been designed by a private company but is deployed by the public sector.⁷⁸

Illiterate opacity arises due to an inability to read computer code or understand statistical methods.⁷⁹ This can become an issue with very simple algorithms as most are not able to understand common programming languages. Even those that can may still struggle to understand what the algorithm is doing. Machine learning models are often based on

⁷⁴ IBM Cloud Education, 'Artificial Intelligence (AI)' (2020) <<https://www.ibm.com/uk-en/cloud/learn/what-is-artificial-intelligence>> accessed 13 May 2022

⁷⁵ Tal Zarsky, 'The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making' (2016) *Science, Technology, & Human Values*, 41(1), 119

⁷⁶ *Ibid*

⁷⁷ Jennifer Cobbe (n 13)

⁷⁸ *R (Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058, [196]

⁷⁹ Jennifer Cobbe (n 13)

complex mathematical theories, making them difficult to understand.⁸⁰ This issue goes further when we consider complex machine learning and deep learning algorithms. These are often so complex that it is impossible for a human brain to understand or explain their reasoning. Deep learning takes this a step further as it has many hidden layers to its neural network making it impossible to know how the data is being manipulated or weighted by the algorithm.⁸¹ This 'black box' nature of complex algorithms is the final form of algorithmic opacity, intrinsic opacity. For these reasons, algorithm developers are often faced with a choice between transparency and accuracy. It is often the case that the more datapoints an algorithm considers and the more complex its manipulation of the data is, the more accurate the algorithm becomes.⁸² When trying to address illiterate and intrinsic opacity, it is often impossible without losing some of the predictive power of the algorithm, creating more issues that it solves.

There have been criticisms aimed at the developers of algorithms for concealing their algorithms or the data used to train them.⁸³ There have also been suggestions companies could be doing more to increase the transparency or readability of their algorithms.⁸⁴ However, there are legitimate reasons beyond selfish commercial interests why a company may wish to protect details about their algorithms. Public knowledge of an algorithm's considerations and the weightings it places on information creates the potential for the algorithm to be exploited through 'gamification'.⁸⁵ This is where individuals change the information given to the algorithm in order to gain more favourable outcomes.⁸⁶

3 Discussion on public bodies

3.1 Public Sector Equality Duty

Unlike the private sector, the public sector has taken a more socially responsible approach

⁸⁰ Ibid

⁸¹ Ibid, 641

⁸² Pragma Paudyal & William Wong, 'Algorithmic Opacity: Making Algorithmic Processes Transparent through Abstraction Hierarchy' (2018) Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62(1), 192

⁸³ Cathy O'Neil (n 14)

⁸⁴ Lorna McGregor et al, 'International human rights law as a framework for algorithmic accountability' (2019) International & Comparative Law Quarterly, 68(2), 309-343

⁸⁵ Bruno Lepri et al, 'Fair, Transparent, and Accountable Algorithmic Decision-making Processes' (2018) Philosophy & Technology, 31(1), 611, 621

⁸⁶ Ibid

to ADM.⁸⁷ The Public Sector Equality Duty (PSED) has created a positive obligation on public bodies to have “due regard” to the need to eliminate discrimination.⁸⁸ However, it was initially unclear what would be required to ensure compliance with that criterion. Who should bear the ultimate responsibility of testing for and proving discrimination by an algorithm, the public body, or the public. This question was put before the courts in *R (Bridges) v South Wales Police (SWP)*.⁸⁹ SWP had been running trials using a new live facial recognition technology. A camera would pick photos of passers-by and, using an algorithm; it would compare these images to a watchlist compiled by SWP, notifying officers if it found a match.⁹⁰ SWP had conducted an impact assessment prior to its deployment.⁹¹ In it, SWP considered whether there was the possibility for direct discrimination because of the technology. However, Bridges, a civil liberties campaigner, challenged SWP's use of the technology on multiple grounds, including human rights and data protection.⁹² Bridges challenged SWP's actions as a failure of their duties under PSED.⁹³ SWP had prepared an equality impact assessment prior to the facial recognition trial however it had only considered whether the technology may be directly discriminatory.⁹⁴ SWP had not considered any possible indirect discrimination. In Bridges' view, this was unacceptable. The divisional court rejected this argument in 2019.⁹⁵ It ruled that Bridges' expert evidence had only suggested a general risk of discrimination and not any “specific comment” about the risk.⁹⁶ Furthermore, the divisional court found that Bridges had failed to establish any specific reason to suspect that the algorithm may be worse at classifying women and people of colour.⁹⁷

Bridges appealed this decision in 2020.⁹⁸ The court took a much tougher approach to PSED than the divisional court. It ruled that PSED required SWP to take proactive steps to gather information on the impact of the facial recognition technology on those groups with protected characteristics.⁹⁹ This was in contradiction to the divisional court where they rejected Bridges case on the basis that he could not prove discrimination. The police had attempted to argue

⁸⁷ Jack Maxwell & Joe Tomlinson, ‘Proving algorithmic discrimination in government decision-making’ (2020) *Oxford University Commonwealth Law*, 20(2), 352-356

⁸⁸ Equality Act 2010, s 1(1)

⁸⁹ *R. (on the application of Bridges) v Chief Constable of South Wales* [2020] EWCA Civ 1058, [2020] 1 W.L.R. 5037b

⁹⁰ *R. (on the application of Bridges) v Chief Constable of South Wales* [2019] EWHC 2341, [2020] 1 W.L.R. 672, [20] – [42]

⁹¹ *Ibid* [20]

⁹² *Ibid* [6]

⁹³ (n 89), [32]

⁹⁴ *Ibid* [53]

⁹⁵ (n 90) [159]

⁹⁶ *Ibid* [153]

⁹⁷ *Ibid* [154]

⁹⁸ (n 89)

⁹⁹ *Ibid* [51]

that their use of a human failsafe meant that they had satisfied their PSED requirements.¹⁰⁰ The court disagreed, suggesting that the use of a failsafe did not discharge SWP of their procedural obligation to assess the technology for bias.¹⁰¹ Furthermore, the court also questioned whether a human failsafe could be trusted not to make mistakes.

The court also found that the police officer conducting the analysis of the system prior to its deployment was not adequately qualified to assess the technical aspects of the system; illiterate opacity.¹⁰² Much of the technical analysis of the system came from the system's manufacturer. The manufacturer had given assurances that the system was unbiased and that it had been trained on a variety of faces from both sexes and a wide range of ethnic minorities.¹⁰³ However, Bridges argued that it was impossible to assess the algorithm for a discriminatory impact without access to the training data. This was something that the manufacturer refused to give on the grounds of commercial confidentiality.¹⁰⁴ The court concluded that the police could not rely on general assurances from the manufactures to discharge PSED especially as it was a non-delegable duty.¹⁰⁵

The Court of Appeal's decision in Bridges has been regularly cited as evidence that the PSED has allowed the public sector to lead the way in its approach to the responsible adoption of ADM processes. The court's judgment means that in future, public bodies must do everything reasonable to ensure that their use of ADM does not have the potential for discriminatory outcomes.¹⁰⁶ This moved away from the low bar set by the divisional court. The previous judgement allowed public bodies to ignore potential risks by relying upon the current lack of evidence of discrimination as a sufficient discharge of PSED.

The judgment has also raised the bar as to the analysis necessary to discharge PSED and confirmed that this cannot be outsourced to private bodies. Lastly, the judgment has ended the debate surrounding the effectiveness of human a failsafe when using ADM in the public sector. The court ruled that the use of a human failsafe was insufficient to discharge the PSED as it is a duty to the process that must be followed and not the outcome.¹⁰⁷ The court did not attempt to tackle the effectiveness of such a failsafe.

It is important to note that this decision was solely focused on the due diligence required

¹⁰⁰ Ibid [184]

¹⁰¹ Ibid [185]

¹⁰² Ibid [191]

¹⁰³ Ibid [199]

¹⁰⁴ Ibid [199]

¹⁰⁵ Ibid [199]

¹⁰⁶ Jack Maxwell & Joe Tomlinson (n 87)

¹⁰⁷ R (Bridges) v Chief Constable of South Wales [2020] (n 89), [184]

under PSED, before deploying an ADM process. It has not made any fundamental changes to the way ADM is used or the remedies available to those that have been victims of a discriminatory decision. While it is undoubtedly an improvement on the previous position on the use of ADM, it has done little to address the plethora of concerns about the growing prevalence of ADM in the public sector. Unfortunately, discriminatory algorithms will still make it through the PSED screening. Those that do will fall into two categories, algorithms that make decisions that have a legal or 'significant affect' on the data subject and those that do not.

3.2 Use of ADM and DPIA

Article 22 of the UK GDPR restricts the making of decisions using personal data by processes without human involvement.¹⁰⁸ Especially where this decision will produce a legal or similarly significant effect. There are a few situations where such decisions may be made. It being necessary for the performance of a contract or based upon the explicit consent of the data subject being the most common.¹⁰⁹ However, unlike most private bodies, many of the functions of the public sector are monopolistic in their nature. Moreover, often the refusal to engage with the bureaucratic process will result in the denial of an essential service with no alternative options. For these reasons, it cannot be said that individuals can freely give consent to be the subject of an automated process. Recital 43 of the EU GDPR makes this clear by saying that, "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".¹¹⁰ The recital goes on to specify public authorities as an example of a data controller that would struggle to gain consent.

UK GDPR requires organisations to perform a Data Protection Impact Assessment (DPIA) whenever they are planning to perform data processing that is "likely to result in a high risk to individuals".¹¹¹ This includes specific types of 'high risk' processing such as using ADM to make significant decisions about individuals, processing special category data or when deploying innovative technology.¹¹² This means that in the UK, any data processor wishing to carry out ADM which has a clear potential to create discriminatory consequences, is

¹⁰⁸ UK GDPR, Art 22

¹⁰⁹ Information Commissioners Office, 'Rights related to automated decision making including profiling' (2022) <<https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling>> accessed 7 May 2022

¹¹⁰ EU GDPR, Recital 43

¹¹¹ UK GDPR, Article 35 (1)

¹¹² Ibid, Article 35 (3) & (4)

required to carry out a DPIA. There have been suggestions that this requirement alone should prevent discrimination by ADM from taking place.¹¹³ In conducting a DPIA, organisations are required to carry out an analysis of the risk of unfair or illegal discrimination. However, recent cases have shown that this analysis is often inadequate in detecting discriminatory practices.¹¹⁴ In *R v Bridges*, the court found that the DPIA conducted by South Wales police was inadequate for a number of reasons.¹¹⁵ It had not considered relevant human rights implications for all those that may be affected, and importantly, the police's analysis of the algorithms potential for discrimination was unsatisfactory.¹¹⁶ The officer in charge of the analysis had little experience with such complex algorithms and had not had access to all the training data as this had been withheld by the algorithm's developer.¹¹⁷

3.3 Automation Bias

In order to circumvent the restrictions placed on public bodies when using ADM, organisations may rely on a human failsafe so that any decisions made with the aid of ADM are not made solely on automated processes and therefore do not fall under Article 22.¹¹⁸ However, there are many well-documented issues with the ability of humans to oversee ADM processes. According to advice published by the European Article 29 Data Protection Working Party, human oversight must be meaningful and not just a 'token gesture'.¹¹⁹ This means that it is not enough to simply have a human apply a decision without further analysis.

This raises several issues for the use of ADM in the public sector. Firstly, humans have a well-documented tendency to over-trust computerised aids.¹²⁰ This phenomenon is known as automation bias. ADM is often a fairer and more objective decision maker than their human counterparts, able to process more information and create more reliable results.¹²¹ However, this, along with the human tendency to find a path of least cognitive resistance,

¹¹³ Frederik Borgesius (n 60); Bryce Goodman, 'A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection' (2016) 29th Conference on Neural Information Processing Systems

¹¹⁴ Frederik Borgesius (n 60)

¹¹⁵ *R (Bridges) v Chief Constable of South Wales* [2020] (n 89) [145]

¹¹⁶ *Ibid*

¹¹⁷ *Ibid* [199]

¹¹⁸ UK GDPR, Art 22

¹¹⁹ Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation' (2017) 2016/679, 10

¹²⁰ Kate Goddard et al, 'Automation bias: a systematic review of frequency, effect mediators, and mitigators' (2012) *Journal of the American Medical Informatics Association*, 19(1), 121-127

¹²¹ Centre for Data Ethics and Innovation (n 2), 1

can cause humans to accept results produced by ADM that are clearly implausible.¹²² This is what Daniel Kahneman calls system one thinking.¹²³ Research has found that people will place a similar weight on a computer-generated decision as they would if a fellow team member had made the suggestion.¹²⁴ This clearly raises doubt as to the ability of human failsafe to provide effective discretion. There is currently a lot of research looking at methods to reduce this reliance on automated decisions.¹²⁵ Suggestions such as giving an automated decision as a likelihood percentage rather than a definitive decision has been shown to reduce this effect but by no means eliminate it.¹²⁶

A second issue arises after a decision has been made. Due to algorithmic opacity, ADM has an extremely limited ability to give reasoning for its decisions. However, there are situations where reasoning must be given or can be requested.¹²⁷ Although this is not a general rule for public bodies, circumstances where reasoning is not needed are rare.¹²⁸ Furthermore, if a decision is contested through judicial review, the court will require the reasoning behind the decision. Therefore, the government legal service recommends recording the reasoning for decisions at the time they were made and offers advice on how this should be done.¹²⁹ This is an issue for decisions made with the aid of ADM as the decision maker will be unaware of the reasoning behind the algorithm's advice. By acting upon the decision made by the algorithm the decision maker should record a reason. They cannot use the advice of the algorithm as a reason as this would undermine the function of the human failsafe, causing the decision to be prohibited by section 49 of the DPA 2018.¹³⁰ Cobbe argues that this would cause a human failsafe to construct reasoning that reaches the same conclusion as the Algorithm after the decision has been made.¹³¹ This is a manifestation of 'choice supportive bias, another well documented human tendency to retrospectively ascribe positive attributes

¹²² Rita Gsenger & Toma Strle, 'Trust, Automation Bias and Aversion: Algorithmic Decision-Making in the Context of Credit Scoring' (2021) *Interdisciplinary Description of Complex Systems*, 19(4), 540-558

¹²³ Daniel Kahneman, 'Of 2 Minds: How Fast and Slow Thinking Shape Perception and Choice' (2012, *Scientific America*) <<https://www.scientificamerican.com/article/kahneman-excerpt-thinking-fast-and-slow>> accessed 5 August 2022

¹²⁴ Rita Gsenger & Toma Strle (n 122), 545

¹²⁵ Kate Goddard (n 120)

¹²⁶ *Ibid*

¹²⁷ Government Legal Department, 'The judge over your shoulder — a guide to good decision making' (2018)

<https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746170/JOYS-OCT-2018.pdf> accessed 23 May 2022, 56

¹²⁸ *Ibid*, 55

¹²⁹ *Ibid*

¹³⁰ Data Protection Act 2018, s 49

¹³¹ Jennifer Cobbe (n 13)

to a decision an individual had made.¹³²

3.4 Lack of Reasoning

There is also increasing concern that the inability of ADM to give reasoning for its outputs will leave public bodies struggling to defend discrimination claims.¹³³ Furthermore, this lack of reasoning is also contrary to expected practices in administrative decision-making. Despite decades of campaigning by legal scholars, there is still no common law duty for public bodies to provide reasoning for their decisions.¹³⁴ However, circumstances where reasoning is not required are rare.¹³⁵ Moreover, recording and providing reasoning behind significant decisions by public bodies is still usual practice and highly recommended.¹³⁶ This is for multiple reasons; often, there is a precedent of reasons being provided for a decision and therefore it would be a failure of a “legitimate expectation” not to continue to provide such reasoning.¹³⁷ Other situations where reasoning is required are where a decision is inconsistent with previous policy or where the matter is of such importance that the principle of fairness requires reasoning to be given.¹³⁸ The government’s legal department notes that the prevalence of judicial review means that fairness requires reasoning to be given in most cases. This means that across the 119 ‘reasons’ challenges against public bodies between 2014 and 2019, almost all of them (95%) concerned the inadequacy of the reasoning rather than a lack of reasoning all together.¹³⁹ However, it is expected the increase in the use of ADM will exacerbate these issues.

Cases such as *R v Higher Education Funding Council* have shown that public bodies can often avoid providing reasoning for a decision if it would be particularly onerous or create an unreasonably high administrative burden.¹⁴⁰ This is the likely defence for a public body accused of failing to adequately reason its decision.¹⁴¹ ADM does not produce reasoning itself and while in some circumstances, a highly skilled expert may be able to interrogate an

¹³² Mahdi Kafaee et al, ‘Choice-supportive bias in science: Explanation and mitigation’ (2021) *Accountability in Research*, 28(8), 528-543

¹³³ Jennifer Cobbe (n 13)

¹³⁴ Government Legal Department (n 127), 55

¹³⁵ *Ibid*, 56

¹³⁶ *Ibid*

¹³⁷ *Ibid*, 53-55

¹³⁸ *Ibid*, 48

¹³⁹ Joanna Bell, ‘Reason-Giving in Administrative Law: Where are We and Why have the Courts not Embraced the ‘General Common Law Duty to Give Reasons’?’ (2019) *Modern Law Review*, 82(6), 983-1008

¹⁴⁰ *R v Higher Education Funding Council, ex p Institute of Dental Surgery* [1994] 1 All ER 651, [665]–[666].

¹⁴¹ Jennifer Cobbe (n 13), 348

algorithm, this would create an exceedingly high administrative burden. As a result, it is argued that less and less thought will be put into creating a reasoning for each decision as it is likely unnecessary.¹⁴² Where a reasoning is given by a human failsafe, this is likely to be insufficient due to the biases discussed above. However, Bell argues that those concerned about the lack of a general duty to give a reason have missed the ever-rising number of specific statutory duties to give reasoning for a decision.¹⁴³ Indeed, over half of the 119 cases mentioned above were concerning such a statutory duty.¹⁴⁴ Furthermore, most decisions producing legal or similar effects on the subject will require detailed reasoning to be recorded in the name of fairness.

When considering allegations of discrimination, the courts will require those decision makers to produce a coherent line of reasoning explaining why a decision was made without the reliance on protected characteristics. If a public body is unable to produce an explanation for the decision that was recorded at the time, this will only serve to undermine a defence.¹⁴⁵ For this reason, it will be imperative for those making decisions with the aid of ADM to critically consider the reasoning behind a decision. Reliance on the output of ADM without independent justification will likely serve as evidence that the human failsafe has not made a meaningful contribution to the decision and therefore the decision will be unlawful.

3.5 Data Minimisation and Proxies

A final issue for those using ADM in the public sector is the restrictions on what factors may be considered when making a decision. GDPR and administrative law require that those making decisions may only consider that which is necessary for the decision to be made.¹⁴⁶ This is the GDPR principle of 'data minimisation'.¹⁴⁷ A superficial understanding of ADM could leave one with the conclusion that the use of ADM would give public bodies a new tool to prevent irrelevant information, protected characteristics, or conflicting interests from being involved in decision-making. The logic is that if something is not relevant or should not be considered, simply do not give the information to the algorithm and instantly, you have a more impartial judge than any human standard could hope to achieve. However, machine

¹⁴² Ibid

¹⁴³ Joanna Bell (n 139), 997

¹⁴⁴ Ibid, 985-988

¹⁴⁵ Ibid, 991

¹⁴⁶ Jennifer Cobbe (n 13), 350-353.

¹⁴⁷ UK GDPR, Article 5 (1)(c)

learning and, especially deep learning have the ability to create proxies for data withheld from it that may be relevant. A proxy is a factor that an algorithm has learnt has a strong correlation with the occurrence of another, withheld factor.¹⁴⁸ This can cause it to place undue weight on a seemingly irrelevant factor. If such a withheld factor is a protected characteristic than this will likely constitute indirect discrimination. Unfortunately, it is near impossible to determine why an algorithm has decided to place certain weightings on certain factors, so proving this would be difficult.¹⁴⁹ While public law does ensure that irrelevant or protected factors are not included in decision-making, there is no restriction or guidance on the weighting that may be placed on a particular factor. This creates the current situation where it is impossible to know what information is truly being considered by an ADM process.¹⁵⁰

3.6 Proving Discrimination

Automated decision making has and will likely continue to prove an invaluable tool in government. However, it does have the potential to produce undesirable or discriminatory results if not carefully implemented or monitored. While data protection law does place restrictions on the use of ADM, there is overwhelming evidence that these restrictions do not neutralise the threat that ADM poses.¹⁵¹ There is a wealth of research suggesting that GDPR's aims of restricting unsupervised ADM are largely ineffective due to the fallibility of human oversight.¹⁵² Furthermore, GDPR's restrictions on processing special category data has not only had the unintentional effect of reducing the willingness of organisations to proactively seek out discrimination but has also compounded the issue of transparency in ADM.¹⁵³ Much of the special category data may still have predictive power due to, systemic bias or discrimination in the data collected. This could allow an algorithm to create proxies and in doing so, add an additional layer of complexity to its reasoning, making interrogating such an algorithm even harder.¹⁵⁴ It may also lure the algorithms operator into a false sense of security, reducing the scrutinization of its outputs.

However, recent developments in the law will force those using ADM to carry out more

¹⁴⁸ Betsy Williams et al, 'How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications' (2018) *Journal of Information Policy*, 8(1), 78-115.

¹⁴⁹ *Ibid*, 101

¹⁵⁰ Tal Zarsky, 'The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making' (2015) *Science, Technology, & Human Values*, 41(1), 118, 127.

¹⁵¹ Jennifer Cobbe (n 13)

¹⁵² Rita Gsenger & Toma Strle (n 122)

¹⁵³ Frederik Borgesius (n 60)

¹⁵⁴ *Ibid*

expensive impact assessments.¹⁵⁵ This is especially true in the public sector where PSED will force public bodies to be incredibly proactive in their analysis and ongoing assessment of the impact their use of ADM may have. However, this will not address many of the concerns raised regarding the opacity of ADM. Algorithms, especially complex ones, are incredibly difficult to understand and are sometimes incomprehensible. This reduces the availability of reasoning and transparency to those affected by the decisions of such algorithms. However, there is hope in that it is not necessary to know the process behind a decision to effectively bring a discrimination claim, providing that a historical pattern of discrimination can be proved.¹⁵⁶

This difficulty in bringing a discrimination claim of any kind was recognised by the Equality Act 2010. Section 136 deals with the burdens of proof.¹⁵⁷ It puts the burden on the claimant to prove that, in the absence of another explanation, discrimination has occurred.¹⁵⁸ The burden then shifts to the respondent to provide an explanation. This two-stage approach to the burden of proof was confirmed in *Royal Mail Group Ltd v Efofi*.¹⁵⁹ This means that the key threshold for those bringing a claim is to show that, without another explanation, the facts infer discrimination.¹⁶⁰ This test cannot be passed by merely showing that A, with a protected characteristic, has been treated less favourably than B. The claimant must show that there was discrimination or that there could have been. However, if this pattern can be shown to be widespread, then a claim may be successful.¹⁶¹ If a claimant can prove that the output of an algorithm consistently favours those of a certain, protected group. In this regard this type of claim would be similar to that of *Rihal v London Borough of Ealing*.¹⁶² Here the court accepted evidence of a 'glass ceiling' within the housing department based on the proportion of non-white managers compared to other departments within the council. Similarly, it would be enough to infer discrimination if the training data used can be shown to be biased as the resulting model would likely reflect this. It should fall on the respondent to provide an alternative explanation to show how they have prevented this.

ADM is more likely to create indirect discrimination as it is easier to remove a protected characteristic from consideration than ensure that the output is not disadvantaging a group.

¹⁵⁵ *R (Bridges) v Chief Constable of South Wales* [2020] (n 89)

¹⁵⁶ Joe Atkinson, 'Automated management, digital discrimination, and the Equality Act 2010' (2020) *Employment Law Bulletin*, 159, 3-6.

¹⁵⁷ Equality Act 2010, s 136

¹⁵⁸ *Ibid*, s 136 (2)

¹⁵⁹ *Royal Mail Group Ltd v Efofi* [2019] EWCA Civ 18

¹⁶⁰ *Madarassy v Nomura International plc* [2007] EWCA Civ 33

¹⁶¹ Joe Atkinson (n 156)

¹⁶² *Rihal v London Borough of Ealing* [2004] EWCA Civ 623

Because of this, those wishing to bring a claim may be able to sidestep much of the consideration surrounding the opacity of the algorithm as they will not need to show why a provision, criterion, or practice is producing a discriminatory effect, only that it does.¹⁶³

In rebutting discrimination claims, those using ADM may find the opacity of ADM will create issues in creating more innocent explanations for seemingly discriminatory outcomes. As a result, this two-step approach to the burden of proof in discrimination cases does not help diminish many of the concerns relating to opacity in ADM as this may still prove to be a barrier to the truth and a just outcome.

4 REFORM

The use of ADM has a well-established potential to create unfair or unfounded decisions.¹⁶⁴ When these decisions are consistently biased against a protected characteristic, then the decision is prohibited by the Equality Act.¹⁶⁵ However, detecting discriminatory decisions can be difficult. Especially when those operating ADM do not collect data on the output and the protected characteristics of the data subject. This is often due to the burden of responsibility placed on those that collect and process such data.¹⁶⁶ When considering reform, a proposal to make the collection of such data compulsory would help detect unsuitable ADM. However, this could create issues such as increased administrative load or, importantly, an increase in breaches of highly personal data. Unintended consequences such as these, make suggesting reform difficult as addressing one issue could create another, potentially worse issue.

To assess what reforms would help reduce opacity in ADM and its potential to discriminate, it is important to look at how other legal systems are attempting to tackle the problems ADM can create. In Canada, a recent supreme court judgment has created obligations on public bodies using ADM to ensure that the algorithm has been assessed for accuracy specifically for each ethnicity the ADM is considering. In *Ewert v Canada*, the court considered a psychological assessment tool used to assess the risk of recidivism of inmates, to help advice on release dates.¹⁶⁷ Ewert, of indigenous Métis decent, argued that the Correctional Service of Canada (CSC) had not fulfilled its obligation to ensure that the data used in the

¹⁶³ *Essop v Home Office* [2017] UKSC 27, [9] - [11]

¹⁶⁴ Joe Atkinson (n 156)

¹⁶⁵ Equality Act 2010

¹⁶⁶ Frederik Borgesius (n 60)

¹⁶⁷ *Ewert v Canada* [2018] SCC 30

decision making was “as accurate, up to date and complete as possible”.¹⁶⁸ Ewert argued that the algorithm had not been satisfactorily tested on those that shared his ethnicity. As a result, the CSC could not rely on its testing of the algorithm to ensure it was reliable in its output.¹⁶⁹ The court agreed with Ewert, creating a new requirement for decision makers in Canada to consider the profiles of those that will be subject to their ADM. This should reduce the chances of discrimination through ADM due to badly applied data.

Professor Borgesius, in a research paper for the Council of Europe, has proposed several other potential ways to reduce the chances of unchecked discrimination in ADM.¹⁷⁰ Along with better education surrounding the issues, especially automation bias, he has suggested that all newly deployed ADM processes be subject to a sunset clause.¹⁷¹ This would mean that after a set amount of time, one year as an example, the outputs are scrutinised by an independent statistics body such as the UK’s Office of national statistics. While this would likely result in more discriminatory ADM processes being detected, it is not a perfect solution. To provide enough data to the independent statistics body, special category data would have to be collected for all data subjects, regardless of the relevance to the decision being made. Nonetheless, a second independent review of active ADM processes using real data would likely reduce the risk of long-term discrimination.

As a result of its newfound freedoms post Brexit, the UK government has issued a consultation paper looking at the ways it needs to reform the UK’s GDPR to unleash the UK’s potential as a science and technology superpower. In the paper titled “Data: A New Direction” the government discusses the limitations of the current law and looks at ways that this could be reformed.¹⁷² The consultation highlights many of the issues discussed above. It argues that the wording of Article 22, especially ‘similarly significant effects’ and ‘solely automated’ does not provide adequate guidance for data processors.¹⁷³ The paper highlights how the wording has caused data processors to be overly cautious in its application of ADM and points to how this has stifled innovation.¹⁷⁴ It has also recognised how ‘solely automated’ has the potential to make data processors circumvent this requirement by placing a token human

¹⁶⁸ Corrections and Conditional Release Act (1992, c.20), s 2

¹⁶⁹ Ewert v Canada (n 163), [2]

¹⁷⁰ Frederik Borgesius, ‘discrimination in artificial intelligence, and algorithmic decision-making’ (2018, Council of Europe) <<https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>> accessed 8 August 2022

¹⁷¹ Ibid, 54

¹⁷² Department for Digital, Culture, Media & Sport, ‘Data: a new direction’ (2021) <<https://www.gov.uk/government/consultations/data-a-new-direction>> accessed 23 May 2022

¹⁷³ Ibid [95] – [98]; UK GDPR, Article 22

¹⁷⁴ Ibid [97]

in the decision-making process.¹⁷⁵

The consultation paper also takes issue with the GDPR requirement for human oversight arguing that it will not be 'practical and proportionate'.¹⁷⁶ As a result, the Taskforce on Innovation, Growth and Regulatory Reform has suggested that the government remove Article 22 to allow for increased use of solely automated decision making when it is in a legitimate or public interest.¹⁷⁷ This would undoubtedly increase the prevalence of discrimination by ADM.

Conclusion

To conclude, it is the UK's governments stated objective to increase the prevalence of ADM in the coming years. ADM promises a faster, cheaper, and more reliable alternative to human decision making. However, there are concerns that the increased use of ADM may reduce transparency and equality in government decision making.¹⁷⁸ Computer algorithms can inherit human biases from badly formed data sets or perpetuate existing inequalities in a negative feedback loop.¹⁷⁹ The computer science profession is making extraordinary strides in reducing these risks and the extraordinary work of equality campaigners is bringing the issues of bad data collection and systemic bias into the minds of data processors in all professions.¹⁸⁰ Unfortunately, there is considerable work to be done and the use of ADM today still poses a real risk of inadvertently creating discriminatory decisions. This is most concerning in the public sector as governments have immense power to grant support or restrict liberty.¹⁸¹ Furthermore, their monopolistic nature means that there is rarely the option of an alternative.

Over the past century, England and Wales have been slowly developing the law to better protect those that have been wronged by discriminatory decisions. However, these developments have been made to combat discriminatory decisions made by humans. The idea that anything nonhuman could serve as a useful decision-maker would have been laughable until very recently. Because of this, there has been considerable concern that the

¹⁷⁵ Ibid [97]

¹⁷⁶ Ibid [100]

¹⁷⁷ Iain Smith et al, 'Taskforce on Innovation, Growth and Regulatory Reform' (2021) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/994125/FINAL_TIGRR_REPORT__1_.pdf> accessed 23 May 2022, [223] – [225]

¹⁷⁸ Jennifer Cobbe (n 13)

¹⁷⁹ Cathy O'Neil (n 14)

¹⁸⁰ Bruno Lepri et al (n 85)

¹⁸¹ European Commission (n 68)

increased use of ADM may result in more discriminatory decisions being made.¹⁸² Not because ADM is more likely to make discriminatory decisions, the opposite is likely to be true, but because ADM struggles to provide a reason for its decisions, the fear is that this will cause discriminatory decisions to go undetected.¹⁸³ Furthermore, this opacity may create a barrier to those trying to prove that they have been the victim of discrimination by ADM. This transparency barriers comes in three forms: illiterate opacity, where opacity comes from an inability to read computer code; intentional opacity, where the algorithms logic is intentionally concealed to protect trade secrets or to reduce the risk of gamification; finally, intrinsic opacity, where the algorithms logic is so complex that human minds are unable to understand it.¹⁸⁴ This final form is the most common, especially with artificial intelligence or neural networks.

This opacity affects the legal framework surrounding equality in the public sector in several ways. As the use of ADM requires personal data to be processed, a Data Protection Impact Assessment (DPIA) is required before ADM can be deployed by the public sector. In this assessment, the public body must assess the risk and take steps to minimise the potential for discrimination to occur as a result. As ADM was first being adopted by the public sector it was clear that these DPIAs were not satisfactorily assessing the risk of discrimination. This led to the *Bridges v SWP* case.¹⁸⁵ Here the courts made a clear outline of the requirements for a DPIA when using ADM. This was a drastic strengthening of the law. Requiring a much tougher impact assessment as well as dispelling the idea that a human failsafe would serve to absolve public bodies of many of their obligations.¹⁸⁶ This strengthening of the pre-deployment requirements has removed many of the legitimate concerns surrounding the use of ADM. As a result of *Bridges*, public bodies will extensively test new ADM systems for bias prior to deployment. The law could be developed further in this area by requiring regular reviews of the outputs to ensure they are equitable.¹⁸⁷ Although there are concerns that this would require additional special category data to be collected, it is common for public bodies to ask for this data for monitoring purposes in many areas such as job applications. It would likely be publicly palatable for them to start requesting demographic data on those subject to ADM process providing the reasoning is well explained.

As identified by a recent government consultation, Article 22 of the UK's GDPR is in need of

¹⁸² Centre for Data Ethics and Innovation (n 2)

¹⁸³ Jennifer Cobbe (n 13)

¹⁸⁴ Jennifer Cobbe (n 13)

¹⁸⁵ *R (Bridges) v Chief Constable of South Wales* [2020] (n 89)

¹⁸⁶ Jennifer Cobbe (n 13)

¹⁸⁷ Frederik Borgesius (n 170)

reform.¹⁸⁸ Its current restrictions on ADM only apply to solely automated decisions, and as a result, decisions made with the aid of ADM have no such restrictions. This is an area of serious concern, given the well-established phenomenon of automation bias.¹⁸⁹ The law, in its current form, is forcing many of those using ADM to use it as an aid to a human operator. This is likely to lure those assessing the risks of ADM into a false sense of confidence, by assuming that the human failsafe will detect any unusual or discriminatory outputs. While the PSED does go some way to protecting the public sector from this trap, it is far from a solution and does nothing to protect the private sector. Moreover, using a human failsafe to create reasonings for the decisions of ADM will prove to be a barrier to justice for those entitled to the reasoning for a decision made about them.¹⁹⁰ It remains to be seen how claimants will fair trying to prove choice-supportive bias in the given reasoning for a decision. Automation and choice-supportive bias are both well documented human weaknesses that need to be carefully considered in any reform to the law surrounding ADM. ADM does have the potential to remove such human fallacies from the government, but this must be done with care. The current situation means that human bias is mixed with poorly scrutinised ADM and threatens to undermine public trust in ADM's potential.¹⁹¹

Thankfully, once discrimination has been detected, many of the fears that the opacity of ADM will lead to an inability to prove discrimination are unfounded. Providing the claimant can show a historical pattern of those with a protected characteristic being denied opportunities, then the burden will fall to the public body to provide an explanation for this.¹⁹² This means that the law will not place an extra burden on those seeking to bring a discrimination case concerning a decision made by ADM. However, there is a legitimate concern that the opacity of ADM may serve to undermine the defence of those using it.¹⁹³ The public body may find it difficult to provide an explanation, innocent or otherwise, for the decisions made by ADM. It remains to be seen how big of an issue this becomes and how engaged legislators will be in addressing it.

One further issue with ADM that the law is completely unequipped to deal with is discrimination against non-protected characteristics. So far, discrimination law has only developed to deal with discrimination against protected characteristics. These characteristics have been identified as in need of protection because there has been a history of

¹⁸⁸ Department for Digital, Culture, Media & Sport (n 172)

¹⁸⁹ Kate Goddard (n 120)

¹⁹⁰ Rita Gsenger & Toma Strle (n 122)

¹⁹¹ Department for Digital, Culture, Media & Sport (n 172)

¹⁹² Joe Atkinson (n 156)

¹⁹³ Ibid

discrimination against those with these traits. It is possible the ADM processes start to uncover previously unknown systemic biases or groups that are disadvantaged through no fault of their own. Alternatively, this could arise from anomalies in datasets. This could be anything, such as hair colour, a postcode, an occupation, gym users or a name. As this would not be discrimination as defined by the Equality Act 2010, there would be few available options to remedy this kind of injustice.

Reform to the law in England and Wales seems inevitable. The Information Commissioner's Office and the Centre for Data Ethics and Innovation have both released reports outlining issues with the current law surrounding ADM. The current government's post-Brexit plan also includes unlocking the UK's potential as a science and technology superpower. Both candidates to become the next Prime Minister have also set out plans to remove and reform many of the restrictions inherited from the EU's laws. This is likely to include the GDPR.