IN THE MATTER OF

AUTOMATED DATA PROCESSING IN GOVERNMENT DECISION MAKING

_____

JOINT OPINION

_____

Contents

Introduction

1. The Legal Education Foundation (**"TLEF"**) seeks to identify new and emerging areas of law where there are gaps in the legal analysis and the need for increased understanding. It has recognised that the regulation of the government use of automated processes to make decisions and support human decisions, may be one such new and emerging area of the law. To this end, we have been instructed to consider how the principles of equality law apply in these circumstances.

2. Other legal provisions such as administrative law principles, the Human Rights Act 1998, the Data Protection legislation (which gives effect to the General Data Protection Regulation[1] (GDPR) and Law Enforcement Directive[2]), the Protection of Freedoms Act 2012 (which controls the taking of biometric data and the use of surveillance camera technology) and the Surveillance Camera Code of Practice of June 2013[3] must also be borne in mind and these will be referenced where necessary.

3. Swee Leng Harris, Head of Policy and Public Affairs at TLEF, has highlighted two examples of governmental automated data processing in decision-making that provide specific contexts for us to consider –

   a. The use of automatic checks of DWP and HMRC data to verify residence within the Settled Status application process, and

   b. A risk-based verification (RBV) process utilised by some Local Authorities (LAs) in relation to Housing Benefit and Council Tax Benefit applications.

---

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation): https://publications.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1/language-en.
[2] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2016.119.01.0089.01.ENG.
[3]https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/282774/SurveillanceCameraCodePractice.pdf

ai LAW

CLOISTERS

4. We have structured this paper in the following way -

   a. **Section A**: We map out the role of automated decision-making in government.

   b. **Section B**: We consider the existing equality framework and how it applies to the two case studies identified by Swee Leng Harris.

   c. **Section C**: We consider how the Public Sector Equality Duty applies to automated decision-making undertaken by the government.

   d. **Section D**: We explore the legal avenues available to individuals who wish to challenge automated decision-making via equality law.

   e. **Section E**: We outline the current debate concerning automated decision-making and its governance so as to draw together the issues identified in this paper.

5. Ultimately, we conclude that there is a very real possibility that the current use of governmental automated decision-making is breaching the existing equality law framework in the UK, and is "*hidden*" from sight due to the way in which the technology is being deployed.

6. Notwithstanding these conclusions, we should emphasise that we fully understand the benefits of automated decision-making. In no way should our Opinion be interpreted as endorsing a blanket resistance to technology that has the potential to increase the speed and perhaps accuracy of important government functions. Rather, our Opinion should be read as a caution against the uncritical acceptance and endorsement of automated decision-making because of its potential to cause damaging unlawful discrimination.

A)   Automated decision-making

7. Automated decision-making occurs where a conclusion is reached without any direct, or with only limited human involvement. This automation may simply be the application of a specific, limited, human-created algorithm that is applied to inputted data. The Information Commissioner's Office (ICO) in its guidance on automated decision-making provides an innocuous example of a fully automated

decision-making process based on an algorithm[4] which exemplifies such a simple kind of process -

> An examination board uses an automated system to mark multiple choice exam answer sheets. The system is pre-programmed with the number of correct answers required to achieve pass and distinction marks. The scores are automatically attributed to the candidates based on the number of correct answers and the results are available online.

8. However, automated decision-making can also be undertaken by algorithms using complex Artificial Intelligence (AI) supported by Machine Learning (ML). AI refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.[5] ML is increasingly being used to "*train*" these AI systems. ML has been helpfully described by the International Association of Privacy Professionals (IAPP) as follows[6] -

> **What is machine learning?**
>
> Machine learning is a technique that allows algorithms to extract correlations from data with minimal supervision. The goals of machine learning can be quite varied, but they often involve trying to maximize the accuracy of an algorithm's prediction. In machine learning parlance, a particular algorithm is often called a "model," and these models take data as input and output a particular prediction. For example, the input data could be a customer's shopping history and the output could be products that customer is likely to buy in the future. The model makes accurate predictions by attempting to change its internal parameters — the various ways it combines the input data — to maximize its predictive accuracy. These models may have relatively few

---

[4] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/

[5] This definition is taken directly from the European Commission in its Communication to the European Parliament and other bodies: https://eur-lex.europa.eu/legal content/GA/TXT/?WT.mc_id=Twitter&uri=COM%3A2018%3A237%3AFIN. The Commission goes on to explain that, "*AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).*" For more information, a detailed paper exploring the meaning of AI produced by the European Commission's High-Level Expert Group on Artificial Intelligence: https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

[6] https://iapp.org/news/a/the-privacy-pros-guide-to-explainability-in-machine-learning/

ai LAW

CLOISTERS

> parameters, or they may have millions that interact in complex, unanticipated ways.

9. An example of a complex AI system supported by ML in the public sector is the Harm Assessment Risk Tool (HART) which has been utilised, since 2017, by Durham Constabulary. It deploys an ML algorithm to classify individuals according to their "*risk*" of committing violent or non-violent crimes in the future.[7] This classification is created by examining an individual's age, gender and postcode (which can be a proxy for race).

10. The algorithm used to analyse this information and assign a "*risk*" rating is eye-wateringly complex, with over 4.2 million decision points.[8] The detail of its processing is therefore not comprehensible. This lack of transparency is typical of many AI systems as eloquently described by Jenny Burrell[9] -

> While datasets may be extremely large but possible to comprehend and code may be written with clarity, the interplay between the two in the mechanism of the algorithm is what yields the complexity (and thus opacity).

11. However, as Jennifer Cobbe states in a recent paper, the opacity created by a complex ML algorithm is only one form of algorithmic opacity. She identified three types of opacity as follows[10] -

   a. **Intentional opacity**: This will occur where a system's "*workings*" are deliberately concealed to protect intellectual property.

   b. **Illiterate opacity**: This will occur where a system is "*only understandable to those who can read and write computer code*".

   c. **Intrinsic opacity**: This will occur where ML means that a human will struggle to understand the decision-making process.

---

[7] "*ML Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges*", Alexander Babuta, Marion Oswald and Christine Rinik, RUSI, September 2018: https://rusi.org/sites/default/files/20180329_rusi_newsbrief_vol.38_no.2_babuta_web.pdf
[8] Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes (2018) "*Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality*", Information & Communications Technology Law, 27:2, 223-250, DOI: 10.1080/13600834.2018.1458455: https://doi.org/10.1080/13600834.2018.1458455
[9] Jenna Burrell, "*How the machine 'thinks': Understanding opacity in machine learning algorithms*", 2016: https://journals.sagepub.com/doi/10.1177/2053951715622512
[10] Jennifer Cobbe, "*Administrative law and the machines of government: judicial review of automated public-sector decision-making*", Legal Studies (2019), 1-20.

12. We would add to this list a fourth type of opacity which is one which arises because data sets are deleted due to concerns, driven by the GDPR, that data should only be processed for its collected purpose. A recent example of this form of opacity is evident from the decision of the High Court in *R v The Chief Constable of South Wales & others ex parte Bridges* [2019] EHWC 2341 (*Bridges*) in which Liberty challenged facial recognition technology deployed in Cardiff. The judgment reveals that South Wales police immediately deleted biometric data from its system once the algorithm had concluded that an individual was not a "*match*" for its watch lists. Whilst this policy is laudable in many ways, as we explain later in this Opinion when we discuss the Public Sector Equality Duty, it also creates a potential problem: how can it be assessed whether the system is behaving in a non-discriminatory way if data evidencing its decision-making process is not retained?

13. Importantly, in some situations, these forms of opacity, which may exist simultaneously, might not matter. However, the "*risk*" rating generated by HART is being used by custody officers to make significant decisions concerning people's liberty, for example, whether an individual should be permitted to access an "*out of court*" disposal programme.[11] The lack of any transparency over the use of such highly sensitive data means that accountability for such decisions is rendered much more difficult to achieve. The reasons for the decisions become "*unseen*". The potential for discrimination to occur as a result is an extremely important concern for civil society.

14. The widespread use of facial recognition technology by police forces is another example of government sanctioned algorithms having a direct effect on the public. The Law Society has recently discussed this, saying[12] -

> Facial recognition technologies detect, extract and compare characteristics of faces from images or video against a database of faces to identify potential matches. In a policing context, it can be envisaged that facial recognition be deployed in attempting to detect either suspects, victims, or vulnerable persons. The reasons for doing this can be understood in terms of automation

---

[11] A comprehensive review of the use of this type of "*predictive policing*" technology is outlined in Hannah Couchman's report produced for Liberty entitled, "*Policing by Machine – Predictive Policing and the Threat to Our Rights*":
https://www.libertyhumanrights.org.uk/sites/default/files/LIB%2011%20Predictive%20Policing%20Report%20WE.pdf

[12] "Algorithms in the Criminal Justice System", the Law Society, June 2019:
https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report/

ai LAW

CLOISTERS

and augmentation – facial recognition has been introduced with a view to both more efficiently enable the identification of some individuals, as well as with a view to identify some sought-after individuals who probably would not have become known to police within those contexts otherwise.

Only the first of these applications is in mainstream use today in the justice system, and so will be focused on below.

As of 15 July 2016 there were 16,644,143 images enrolled in the facial image recognition gallery of the Police National Database, searchable using automated facial recognition software. As of that date, the Home Office did not record how many of those faces belonged to individuals not charged of any offence. As of January 2018, that number is reported to have dropped to 12.5m for reasons unknown.

In England and Wales, the highest profile uses of facial recognition technologies in public spaces by the police consist of trials run by the **London Metropolitan Police**, the **South Wales Police** and **Leicestershire Police**. All three forces are trialling technologies produced by NEC, a Japanese firm. The forces have limited ability to oversee or alter the software provided by NEC without the firm taking initiative, with the deputy chief constable of South Wales Police noting that "the tech is given to [them] as a sealed box… [South Wales Police] have no input – whatever it does, it does what it does". This software can run in two main modes: an Identify mode, which functions on pre-recorded images and compares them to a database of held images, which can be in the hundreds-of-thousands, or a Locate mode which works on live-streamed video but against a smaller, pre-filtered database of a few hundreds or thousands. Both approaches provide a list of likely individuals to a staff member who manually examines the results.

Leicestershire Police was one of the first police forces to trial the live Locate facial recognition technology from April 2014, notably using the tool to look for approximately 90,000 'known offenders' at the Download festival in June 2015.

South Wales Police bid to the Home Office Police Transformation Fund in 2016, receiving £1,950,000 over two years (with the force committing £600,000 of their own funds) from January 2017 to deploy automated facial recognition in the context(s) of counter-terrorism; major events; body worn video; mobile phone app(s); automated number plate recognition; and child sexual exploitation. Funding was conditional on an evaluation of the technology being undertaken, which was carried out and published by Cardiff University.

The force deployed the Locate system live at a number of sporting and music events in 2017/18. The threshold for inclusion as an individual of interest was unclear in several of these events to the evaluation team, although was noted to

include strategies targeting pickpockets and mobile phone theft. An initial deployment at the UEFA Champions League saw 2,632 matches, of which only 3% were 'true positives' and only one arrest resulted; later deployments saw the threshold for matching significantly increased, with a small music event, Elvisfest, matching 18 individuals with a 61% true positive rate, and a large boxing match matching 60 individuals with a 9% true positive rate. A newer algorithm deployed later saw true positive rates ranging from 14 to 46%. Newspapers have widely reported on these high false positive rates. Where individuals were approached mistakenly, Cardiff University reported that "for the most part, interactions… were amicable […] operators / officers fully explained the exercise being carried out, and the individuals were invited into the vans to see the software for themselves and to see their own CCTV image alongside the 'match'".'

The force deployed the Identify system on a laptop in the headquarters, with one member of staff a day responsible for facial recognition functions. Initially, many of the images sent from the field were of poor quality, such as mobile snapshots of CCTV, and significant organisational effort was required to train officers to ensure they were only sending high quality images. Concerns with this system were that certain individuals were being repeatedly being matched to many photos, and officers reported that a commonality was that these individuals either had old photographs or had facial disfigurements.

The **London Metropolitan Police** undertook 10 deployments of live facial recognition technology between August 2016 at Notting Hill Carnival and February 2019 in Romford Town Centre. Similarly to South Wales Police, a subset of individuals from the Metropolitan Police Service's databases of photographs were extracted, primarily drawn from photos taken while individuals were in custody but also, controversially, from other sources. An evaluation undertaken by the University of Essex is forthcoming now the trial is complete. However, as with the UEFA Champions League case undertaken by South Wales Police, the London Metropolitan Police has come under heavy fire for the revealed number of false positives, particularly in their Notting Hill Carnival deployment in 2016. The video recordings from the static cameras placed for facial recognition deployment for the London Metropolitan Police are retained for 30 days and then destroyed, with no images extracted. Recognised individuals are retained until one month following the end of the trial, with only the evaluation team provided access.

(Footnotes omitted)

15. If facial recognition technology was always accurate this *might*[13] be less of an issue, but since this is not always the case, its use is troubling, and moreover, the availability and use of these systems is increasing. It is obviously a matter of concern, which has led to the Information Commissioner's Office conducting an investigation whereby it is monitoring the trials carried out by the police. This report is awaited, but the Commissioner has already warned that there is as potential for race and gender discrimination.[14] In the meantime, facial recognition machines and technology are increasingly available on Amazon.[15] News stories, such as those during summer 2019 concerning the use of facial recognition technology in Kings Cross,[16] increasingly relate how many private companies are deploying the technology.

16. Moreover, recent research by Joy Buolamwini and Timnit Gebru in the US has revealed how such technology can have a disparate impact on women and certain racial groups. They pointed out that[17] –

> Recent studies demonstrate that machine learning algorithms can discriminate based on classes like race and gender. In this work, we present an approach to evaluate bias present in automated facial analysis algorithms and datasets with respect to phenotypic subgroups. … We evaluate 3 commercial gender classification systems using our dataset and show that darker skinned females are the most misclassified group (with error rates of up to 34.7%). The maximum error rate for lighter-skinned males is 0.8%. The substantial disparities in the accuracy of classifying darker females, lighter females, darker males, and lighter males in gender classification systems require urgent attention if commercial companies are to build genuinely fair, transparent and accountable facial analysis algorithms.

---

[13] We recognise here, of course, that there are still widely held concerns in relation to privacy and the legitimacy of data processing even where facial recognition technology is accurate.

[14] See the Information Commissioner's blog of the 9 July 2019:
https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/07/blog-live-facial-recognition-technology-data-protection-law-applies/?utm_source=twitter&utm_medium=iconews&utm_term=358cc92e-1e5f-44e7-af5a-cfcb1397e296&utm_content=&utm_campaign=

[15] See, for instance: https://www.amazon.co.uk/s?k=facial+recognition&ref=nb_sb_noss_2

[16] https://www.theguardian.com/uk-news/2019/aug/12/regulator-looking-at-use-of-facial-recognition-at-kings-cross-site

[17] Joy Buolamwini, Timnit Gebru; Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91, 2018, "*Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*":
http://proceedings.mlr.press/v81/buolamwini18a.html

17. Nor are their worrying conclusions unique: independent research published by the University of Essex into the activities of the Metropolitan Police Service (MPS) has also noted that this technology has a poor record of assisting the police in accurately identifying individuals[18] –

> Across the six test deployments MPS officers engaged with 22 individuals as a direct result of a computer generated match judged to be credible by a human operator. 14 of these (63.64%) were verified incorrect matches, eight were verified correct matches (36.36%).

18. That is why, in our view, it is imperative that the use of automated decision-making technology does not pass "*by default*" whether it is being used in a fully automated process or in order to guide or inform the decisions of human actors. As Jennifer Cobbe has pointed out[19] –

> … well-attested psychological phenomenon of automation bias, which means that humans are more likely to trust decisions made by machines than by other people and less likely to exercise meaningful review of or identify problems with automated decisions.

19. This in outline is why, in our view, *all* forms of automated decision-making must be properly analysed by reference to the UK's existing legal framework to see what controls apply to prevent their misuse and to ensure that their use is brought under the Rule of Law. The range of applications is very large and to review all contexts would require a very extensive analysis, so in this paper we shall focus principally on Equality Law.

B)   Equality law and automated decision making

20. In the next paragraphs we outline the relevant equalities legislative framework in the UK.

---

[18] Professor Pete Fussey & Dr Daragh Murray, "*Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*", July 2019: https://48ba3m4eh2bf2sksp43rq8kk-wpengine.netdna-ssl.com/wp-content/uploads/2019/07/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report.pdf

[19] Ibid. Jennifer Cobbe, "*Administrative law and the machines of government: judicial review of automated public-sector decision-making*", Legal Studies (2019), 1-20.

21. At present, no equality legislation within the UK or European law has been specifically crafted to apply to AI and ML. Nonetheless, it is our Opinion that the Equality Act 2010, which applies to service providers in the public sector in Great Britain and the equivalent legislation in Northern Ireland[20] do provide, at least *in theory*, important protection against algorithmic discrimination. The *effective practical application* of these protections is what we shall examine later in this paper.

22. In broad outline, the Equality Act 2010[21] works in four main stages –

    a. First by defining a series of "*Protected Characteristics*"[22] in Chapter 1 of Part 2,

    b. Next, in Chapter 2 of Part 2, by defining what is "*Prohibited Conduct*" in relation to these Protected Characteristics,[23]

    c. Then, in subsequent Chapters and Parts of the Act by defining when such Prohibited Conduct in relation to Protected Characteristics will be unlawful, and,

    d. Finally, by defining the means for obtaining specific remedies for such unlawful action.

23. Prohibited Conduct is made unlawful on a largely, but not an entirely, similar basis for each of the Protected Characteristics. Some exceptions are provided for in

---

[20] Sex Discrimination (NI) Order 1976 https://www.legislation.gov.uk/nisi/1976/1042/contents
Disability Discrimination Act 1995 http://www.legislation.gov.uk/ukpga/1995/50/contents
Race Relations (NI) Order 1997 http://www.legislation.gov.uk/nisi/1997/869/contents/made
Equality Act (Sexual Orientation) Regulations (NI) 2006
http://www.legislation.gov.uk/nisr/2006/439/contents/made) The Northern Ireland Act 1998.
These provisions essentially mirror the Equality Act 2010 and accordingly they are not referred to any further in this advice. We are happy to expand on the Northern Ireland specific legislation if and when necessary.
[21] http://www.legislation.gov.uk/ukpga/2010/15/contents
[22] Chapter 1.
[23] Chapter 2.

relation to specific contexts. Provisions are also made which permit positive action and other specific steps to advance equality.

24. Section 29 of the Equality Act 2010 is the key provision when considering whether governmental acts or omissions of Prohibited Conduct in relation to Protected Characteristics are unlawful. This section applies to individuals and organisations that provide goods, facilities and services or exercise a public function. Such Prohibited Conduct by the state in relation to services and functions that is caused by algorithms and automated decision-making will thus fall to be considered by reference to this section subject to any specific exceptions. Unless those exceptions apply,[24] such conduct will be unlawful and give rise to remedies.

25. Of course, some government services will provide "*welcome*" benefits for people living in the UK, for example, the government has recently used ML to analyse user feedback on government websites and so to improve efficiency.[25] On the other hand, the Equality Act 2010 will, as we explain below, also apply to automated decision-making that has "*unwelcome*" consequences for individuals, such as the HART model utilised by Durham Constabulary to which we refer at paragraph 9 above, and further analyse below.

Protected Characteristics

26. There is a limited class of **Protected Characteristics** which includes age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation. This is in contrast to Article 14 of the European Convention on Human Rights which defines an open-ended list of protected characteristics.

Prohibited Conduct

27. Broadly there are four different kinds of **Prohibited Conduct**: direct discrimination, indirect discrimination, victimisation and harassment. However, in relation to disability, special provisions define Prohibited Conduct more widely, creating obligations to make reasonable adjustments[26] to accommodate disabled

---

[24] And subject to any material time-limits being met.

[25] https://dataingovernment.blog.gov.uk/2016/12/20/using-machine-learning-to-classify-user-comments-on-gov-uk/

[26] See sections 20 – 22 of the Equality Act 2010.

persons and defining a special justifiable disability related discrimination concept.[27] We discuss these briefly below.

28. First, we shall focus on direct and indirect discrimination which are of most relevance to AI and ML, though occasionally harassment might also be relevant. Victimisation claims arise where action is taken to stultify or penalise the exercise of rights and is not directly relevant to this Opinion.

Direct Discrimination

29. **Direct discrimination** is defined by section 13 of the Equality Act 2010 to less favourable treatment. It occurs when someone is treated less favourably because of a Protected Characteristic. Such less favourable treatment can occur in many ways and can arise when a decision-maker is consciously or unconsciously influenced by a Protected Characteristic.

30. The HART model used by Durham Police provides a good example of how algorithms supported by ML might be directly discriminatory since it directly uses information related to the Protected Characteristics of race (since postcodes can be a proxy for race), gender and age. This could easily lead to less favourable treatment on the grounds of Protected Characteristics.

31. Importantly, according to the Supreme Court in *R (on the application of Coll) v Secretary of State for Justice* [2017] UKSC 40, direct discrimination may also occur if:

    a. A rule or provision is applied neutrally to a group, but

    b. Everyone who is disadvantaged by it shares a particular Protected Characteristic, and

    c. Everyone who is advantaged/not disadvantaged does not possess the same Protected Characteristic.[28]

32. Specifically, in *Coll*, the Supreme Court concluded that direct discrimination will arise where the disadvantage experienced by the group relates to the *risk* of a

---

[27] See section 15 of the Equality Act 2010.
[28] See, for example, *R (on the application of Coll) v Secretary of State for Justice* [2017] UKSC 40.

disadvantage occurring, rather than simply the disadvantage having actually materialised.[29]

33. This form of direct discrimination by proxy must be considered carefully in the context of automated decision making since we can see the potential for this type of analysis to be deployed in order to argue that it gives rise to direct discrimination. For instance, it could be argued that poor facial recognition technology is more likely to fail when seeking to identify people with certain Protected Characteristics. In such a case, all people from that racial group would always be at a greater risk of misidentification and accordingly mistreatment in consequence of the use of this technology. In contrast, people without that Protected Characteristic would always be at a lower risk of misidentification. It could be argued following *Coll*, that this risk asymmetry gives rise to such less favourable treatment because of race and as such *prima facie* direct discrimination claim.

34. When such "Prohibited Conduct" is established, it becomes necessary to identify whether the unfavourable treatment is unlawful contrary to the Equality Act 2010. All forms of unfavourable treatment because of a Protected Characteristic are unlawful under the Equality Act other than in relation to direct age discrimination where a service provider always has the potential to show that the less favourable treatment is objectively justified. This means that where age is the relevant Protected Characteristic an extra consideration is in play. In our analysis above, since race is the relevant Protected Characteristic, there would be no scope for the government or other service provider to deploy a justification defence.

35. Following the approach taken in *Coll,* we consider that facial recognition technology could potentially be held to be direct discrimination where it creates different level of risk for different groups. However, as we explain at paragraph 39 below, a different "outcomes based" analysis, would lead to the different conclusion that facial recognition technology gives rise to indirect, rather than a direct, discrimination. We shall explain this concept next.

---

[29] This point is further discussed within the blog by Dee Masters:
https://www.cloisters.com/identifying-direct-discrimination-in-proxy-cases-after-r-on-the-application-of-coll-v-secretary-of-state-for-justice/

Indirect discrimination

36. Section 19 of the Equality Act 2010 defines **Indirect discrimination** as Prohibited Conduct in the following terms –

> (1) A person (A) discriminates against another (B) if A applies to B a provision, criterion or practice which is discriminatory in relation to a relevant protected characteristic of B's.
>
> (2) For the purposes of subsection (1), a **provision, criterion or practice** is discriminatory in relation to a relevant protected characteristic of B's if—
>
> (a) A applies, or would apply, it to persons with whom B does not share the characteristic,
>
> (b) it puts, or would put, persons with whom B shares the characteristic at a particular disadvantage when compared with persons with whom B does not share it,
>
> (c) it puts, or would put, B at that disadvantage, and
>
> (d) A cannot show it to be a proportionate means of achieving a legitimate aim.
>
> (Emphasis added)

37. As we have noted above an algorithm is a logical, that is to say a rules-based, process. That means an algorithm is properly seen as a "*provision, criterion or practice*" (PCP) within the meaning of section 19 (1) of the Equality Act 2010. Although we are not aware of any case law to this effect, we are confident that this analysis is consistent with the relevant Code of Practice, approved by Parliament and issued by the Equality and Human Rights Commission (EHRC) [30] -

> The first stage in establishing indirect discrimination is to identify the relevant provision, criterion or practice. The phrase 'provision, criterion or practice' is not defined by the Act but it should be construed widely so as to include, for example, any formal or informal policies, rules, practices, arrangements, criteria, conditions, prerequisites, qualifications or provisions. A provision, criterion or practice may also include decisions to do something in the future

---

[30] "*Services, public functions and associations: Statutory Code of Practice*", EHRC, at [5.6]:
https://www.equalityhumanrights.com/sites/default/files/servicescode_0.pdf

> such as a policy or criterion that has not yet been applied, as well as a 'one-off' or discretionary decision.

38. Likewise, we consider that computer programs and the basis upon which they analyse and manipulate data, so as to "*train*" algorithms or assist with semi or fully automated decision making, are also PCPs.

39. As we have noted above, the facial recognition technology described above in paragraph 35, could thus also be analysed as potentially being indirectly discriminatory, if the necessary adverse impact was established in relation to a particular Protected Characteristic. For example, women from certain racial groups, might be at a particular disadvantage from its use by being at a greater risk of being mis-identified by facial recognition technology as per section 19 (2) of the Equality Act 2010. If correct, the service provider would only escape liability by being able to objectively justify its system under section 19 (2) (d) of the Equality Act 2010.

Disability related Prohibited Conduct

40. We must mention the special kinds of Prohibited Conduct which apply to the Protected Characteristic of disability alone.

41. Section 15 of the Equality Act 2010 prevents a service provider from treating a disabled claimant **unfavourably for a reason arising from their disability** unless the treatment can be justified. We consider how a section 15 claim might arise in the public sector due to the use of algorithms at paragraph 112 below.

42. Section 20 of the Equality Act 2010 imposes an obligation upon service providers and public authorities to make **reasonable adjustments** in relation to disabled people. This means that where a PCP of the service provider puts a disabled person at a substantial disadvantage in comparison with persons who are not disabled, the service provider must take such steps as it is reasonable to take to avoid the disadvantage.

43. Importantly, the duty to make reasonable adjustments is slightly different for service providers, as opposed to employers, in that it is an *anticipatory* duty i.e. a service provider does not need actual or constructive knowledge that a disabled person will be disadvantaged by the PCP (section 29 (7)). Moreover, service providers are required not simply to avoid any relevant disadvantage created for

disabled people but also, if necessary, to adopt a reasonable alternative method of providing the service or exercising the public function (Schedule 2). There is, however, no requirement to take a step which would fundamentally alter the nature of the service (Schedule 2) nor is there a requirement to take a step which a public authority has no power to take (Schedule 2).

Equality justification

44. It will be noted that there is a common thread running through many of the forms of discrimination made unlawful by the Equality Act 2010, namely the possibility that action that is *prima facie* discriminatory may be justified. This defence is available in relation to direct age discrimination, indirect discrimination on all grounds, and in relation to the special kinds of disability discrimination that we have outlined above.

45. In summary, there are three steps that must be taken to establish a successful justification defence. It must be shown that[31] -

    a. **Legitimate aim**: The measure adopted by the service provider is underpinned by a "*legitimate aim*";

    b. **Effectiveness**: The measure must be capable of achieving that aim; and

    c. **Proportionality**: The measure must be proportionate: it is critical to note that in almost all cases a measure will not be proportionate where the aim could be achieved through a different measure that was less discriminatory or not discriminatory at all.

46. There is some important jurisprudence in relation to these justification arguments. Specifically, it is never possible to justify a practice which is *prima facie* discriminatory on the basis that the aim itself is discriminatory, such as, it is cheaper to discriminate.[32] However, it is possible to justify a practice on the basis of a "*legitimate*" economic aim, such as the creation of a financially viable system in circumstances where resources are limited.[33]

---

[31] See, for example, *R (Elias) v Secretary of State for Defence* [2006] EWCA Civ 1293.

[32] See, for example, *Woodcock v Cumbria Primary Care Trust* [2012] EWCA Civ 330.

[33] See, in an employment context, the cases of C-501/12 *Specht v Land Berlin* and *Heskett v The Secretary of State for Justice* (UKEAT/0149/18/DA).

47. We have to be wary when advising whether, algorithms, AI and ML can be justified within this legal framework since there has been very little direct regulation of the AI sphere, there is as present no case law, and there is no statutory guidance directly addressing this important topic. However, there *is* an extensive world-wide discussion about the ethics of using these tools. We must warn though, that the ethical principles that have so far been discussed (and the standards they seem to entail) are not universally adopted by business, and differ somewhat from country to country, and organisation to organisation. Nonetheless, some agreement is beginning to emerge and cohere. That is why we also believe that this discussion will soon lead to legislative measures or subject specific codes of practice in the UK, and the rest of Europe. Even if that does not happen quickly, we believe a body of jurisprudence will develop that will engage with these discussions about ethics. We fully expect that the way in which the concepts of legitimate aim, effectiveness and proportionality will be applied to automated decision-making, AI and ML, will be materially shaped by the principles and standards that are emerging from these ethical discussions. We explain how this might work in practice below.

Legitimate aim

48. Surprisingly, beyond a clear line of authorities stating that it is never a "*legitimate aim*" to seek to discriminate because it is cheaper, there is a paucity of guidance in the UK's existing case law on what amounts to a "*legitimate aim*" more generally.[34] In truth, it is something of an "*elastic*" and imprecise concept. The passage from the EHRC's Code cited above captures this well stating[35] that a "*legitimate aim*" is one which is "*… legal, should not be discriminatory in itself, and it must represent a real, objective consideration*".

49. There is no doubt that an aim which would involve an illegal act, in the sense that it is not sanctioned by primary or secondary legislation, is not permissible and could never be legitimate. Further, we consider that an aim cannot be a legitimate aim of government if it is clearly unethical. We further consider that the principles and standards that are emerging from these discussions about ethics will provide

---

[34] We know that employers can only discriminate on the grounds of age where the aim is a social policy aim of a public interest nature rather than a purely personal reason relating to the employer's situation such as cost reduction or improving competitiveness since this was established in the Supreme Court case of *Seldon v Clarkson* [2012] ICR 716, *Wright & Jakes* in which we represented Mr Seldon. But, there is no case law or statutory guidance which addresses whether service providers, as opposed to employers, are similarly limited in the terms of what amounts to a "*legitimate aim*".

[35] Ibid. at [5.28].

a starting point for any determinative judicial examination as to whether an aim is "*legitimate*", so it is very important to be aware of them. Many of these ethical principles are set out on our website at www.ai-lawhub.com. We have tried to keep this up-to-date as they evolve and we recommend that readers visit that site to see how matters are developing.

50. One starting point for understanding the state of the discussions on the emerging ethical principles and standards is the UK government endorsed[36] guidance published by the Alan Turing Institute,[37] entitled, "*Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public secto*r". The document recommends working to –

> … a framework of ethical values that Support, Underwrite, and Motivate a responsible data design and use ecosystem. These will be called SUM Values, and they will be composed of four key notions: Respect, Connect, Care, and Protect.

51. Each of these values is worth considering. The "*Protect*" value, in our view, very much fits into the conceptual framework of a "*legitimate aim*". This is how the "*Protect*" value is described within the paper:

> → PROTECT the priorities of social values, justice, and the public interest:
>
> • Treat all individuals equally and protect social equity
>
> • Use digital technologies as an essential support for the protection of fair and equal treatment under the law
>
> • Prioritise social welfare, public interest, and the consideration of the social and ethical impacts of innovation in determining the legitimacy and desirability of AI technologies
>
> • Use AI to empower and to advance the interests and well-being of as many individuals as possible
>
> • Think big-picture about the wider impacts of the AI technologies you are conceiving and developing. Think about the ramifications of their effects and

---

[36] "*Understanding artificial intelligence ethics and safety*":
https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety
[37] https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

> externalities for others around the globe, for future generations, and for the biosphere as a whole

52. Some of the "*Care*" value is also likely to be relevant, though to an extent it is concerned with means as opposed to aims. This principle states –

> → CARE for the wellbeing of each and all:
>
> • Design and deploy AI systems to foster and to cultivate the welfare of all stakeholders whose interests are affected by their use
>
> • Do no harm with these technologies and minimise the risks of their misuse or abuse
>
> • Prioritise the safety and the mental and physical integrity of people when scanning horizons of technological possibility and when conceiving of and deploying AI applications

53. In our Opinion, it is likely that an AI system concerned with social welfare or the public interest would not be held to be legitimate to the extent that the purpose of the system was inconsistent with the aims set out in these "*Protect*" and "*Care*" values.

54. We should also note here that within Europe there is an advanced debate concerning the ethical principles that should govern AI. The leading paper is "*The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)*" which was prepared by the European High-Level Expert Group on Artificial Intelligence (AI HLEG).[38] This independent expert group was set up by the European Commission in June 2018. This paper explains that AI must only be used to improve collective and individual well-being. Another paper, "*The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)*"[39], also produced by the AI HLEG, is important too. In addition to identifying the purpose of "*ethical AI*", it also explains the principles that the AI HLEG consider should define any system which are -

- Respect for human autonomy,
- Prevention of harm,
- Fairness and
- Explicability.

---

[38] https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top
[39] https://ec.europa.eu/futurium/en/ai-alliance-consultation

55. These concepts have sometimes been summarised within a guiding principle that AI should be "*human centric*".

56. It will be immediately apparent that the European approach towards creating ethical AI systems parallels the ideas advanced by the Alan Turing Institute in the UK.   So, notwithstanding Brexit, we anticipate that the European approach may well be relied upon by the judiciary in the UK in terms of understanding "*best practice*" and thinking through questions of proportionality and effectiveness. After all, it is not unusual for judges in the UK to take a comparative approach and to look to other jurisdictions when considering how to develop British law in new and emerging areas.

57. We also consider that, when considering the way in which a justification defence might be formulated, it is important to examine the government's own online guide to AI,[40] which sets out what it considers to be the potential benefits from deploying algorithms in the public sector. This says –

How AI can help

AI can benefit the public sector in a number of ways. For example, it can:

- provide more accurate information, forecasts and predictions leading to better outcomes - for example more accurate medical diagnoses
- produce a positive social impact by using AI to provide solutions for some of the world's most challenging social problems
- simulate complex systems to experiment with different policy options and spot unintended consequences before committing to a measure
- improve public services - for example personalising public services to adapt to individual circumstances
- automate simple, manual tasks which frees staff up to do more interesting work

---

58. These aims are fundamentally concerned with promoting the public interest and align with the two values we have cited.

59. Accordingly, we suspect that the real battleground in any justification defence will be the extent to which the purported legitimate aim is *actually achieved* and / or the

---

[40] https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector

ai LAW

CLOISTERS

extent to which the system is *truly proportionate* (the second and third tests outlined above) as opposed to whether there is a legitimate aim in the first place.

Effectiveness

60. We turn now to the issue of effectiveness. At present, we consider it very likely that some automated decision-making systems would fail this test. For example, if we return to the debate concerning the use of facial recognition technology outlined at paragraphs 9 and 13 above, it will be plain that while it is obviously a "*legitimate aim*" to seek to reduce crime, there are more pressing concerns about the effectiveness of current systems in securing this outcome. If a system that is deployed has a low success rate, producing many false positives or false negatives, its effectiveness must be seriously questioned and for that reason alone it would hardly be justifiable.

Proportionality

61. Finally, we consider proportionality. The Alan Turing Institute's Paper proposes what it calls the FAST Principles that it says should dictate the implementation of any AI system. These actually include two "S"s, thus -

    a. **F**airness.
    b. **A**ccountability.
    c. **S**ustainability.
    d. **S**afety.
    e. **T**ransparency.

62. We consider that these principles will be considered highly material in determining issues of proportionality. We will consider the application of these principles to proportionality questions in closer detail at paragraphs 96 to 105 below. It is sufficient to say at this stage that we predict that there will be situations in which automated decision making, supported by AI and ML, will not be proportionate.

63. With the general equality framework in the UK now mapped out, we turn to consider the two case studies identified by Swee Leng Harris.

64. The Settled Status scheme has been established by the Home Office, in light of Brexit, to regularise the immigration status of European Union (EU), European Economic Area (EEA), Swiss nationals and their families living in the UK.[41] A successful application permits an individual to remain in the UK after 30 June 2021.

65. Settled Status is ordinarily awarded to qualifying individuals who started living in the UK by 31 December 2020 (or by the date the UK leaves the EU without a deal) and who have lived in the UK for a continuous five-year period (known as '*continuous residence*').[42] Five years' continuous residence means that for five years in a row an individual has been in the UK, the Channel Islands or the Isle of Man for at least six months in any twelve-month period. Accordingly, to government statistics released in August 2019, as of 31 July 2019, 1,040,600 EU citizens and their families from across the UK have applied to the EU Settlement Scheme.[43] Overall, the total number of applications that have been concluded, as of 31 July 2019, was 951,700. Of these, 64% were granted settled status and 36% were granted pre-settled status.

66. In order to determine if an individual has been resident for the relevant five year continuous period, where an individual provides their National Insurance number, the Home Office application process uses automated data processing to analyse data from the DWP and the HMRC to verify how long an individual has been in the UK.[44]

67. The benefits data held by the DWP, which is examined by the algorithm, consists of thirteen categories: State Pension and New State Pension, Housing Benefit, Jobseekers Allowance, Employment Support Allowance, Carers Allowance, Universal Credit, Personal Independent Payment, Disability Living Allowance, Income Support, Maternity Allowance, Incapacity Benefit, Attendance Allowance and Severe Disablement Allowance.[45]

---

[41] https://www.gov.uk/settled-status-eu-citizens-families

[42] https://www.gov.uk/settled-status-eu-citizens-families/what-settled-and-presettled-status-means

[43] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/825277/eu-settlement-scheme-statistics-july-2019.pdf

[44] https://www.gov.uk/guidance/eu-settlement-scheme-uk-tax-and-benefits-records-automated-check

[45] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790668/Home_Office_-_DWP_API_EU_Exit_MoU.PDF

68. The precise way in which the automated decision-making algorithm reaches the conclusion that an individual has been resident during certain periods of time is not entirely clear. There is reference to "*a calculation*" in the "*Processing Level Memorandum of Understanding*" between the Home Office and the DWP concerning the Settled Status scheme. [46] The document entitled, "*How the automated check calculates the period of residence*" provides more detail but there is still some opacity-

> If the automated check finds evidence of UK employment or some benefits in a month, it counts as a month of continuous residence. For example, if HMRC records find a monthly pay slip this will evidence one month of continuous residence. If HMRC records find a weekly pay slip, this will also count as a month of continuous residence.
>
> Some records count as evidence of residence for a longer period. For example, an annual self-assessment tax return that contains income from self-employment will count as evidence for that whole tax year, not just the month that the self-assessment return was completed. Find out how the automated check calculates your UK residence.[47]
>
> If the automated check cannot find any evidence of residence for more than 6 months in any 12 month period, this counts as a break in your UK continuous residence. If this happens and you believe you qualify for settled status, we'll tell you immediately during your application which time periods this applies to. You'll then be able to upload photos or scans of documents to your application to show your UK residence for at least 6 months in that 12 month period. You'll only need to provide evidence for the months we specify.

69. A case worker then uses the data provided by the algorithm, which is simply the months that the algorithm has determined the applicant was resident, to "*inform a calculation to determine whether an applicant's UK residence indicates whether they are eligible for consideration*" under the scheme. [48] On 16 July 2019, this system was

---

[46] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790668/Home_Office_-_DWP_API_EU_Exit_MoU.PDF. The equivalent document for the HMRC contains the same information and is available here:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790661/Home_Office_-_HMRC_API_EU_Exit_MoU.PDF

[47] Please note that this link is circular and simply brings the reader back to a website page which links to the document itself.

[48] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790668/Home_Office_-_DWP_API_EU_Exit_MoU.PDF

described in the following way by Caroline Nokes MP[49] to the House of Lords Select Committee on the European Union, Justice Sub-Committee[50] –

> It is important to reflect that these are not automated decisions; every decision goes in front of a caseworker. Although the automated checks that can be done with the DWP and HMRC can obviously stream people's responses very quickly, they still have to be put in front of a decision-maker to verify the criminality and identify checks.

70. It also appears that the case worker will be able to exercise some discretion when reaching a decision but when, how and on what basis that discretion is exercised is unclear.[51]

71. There is also a stage within the process whereby individuals who have provided their National Insurance number, and thereby entered the automated decision-making process, can upload documentation in order to verify periods of residence that could not be confirmed by the algorithm.[52]

72. Importantly, however, an applicant will *not* be informed of the reason that an automated check has concluded that the person does not have continuous residence during a certain period. The rationale for this system is that "… *because doing so may introduce the risk of identity theft and abuse*".[53] Accordingly, another layer of opacity is introduced into the system.

73. Finally, there is a right to seek an administrative review of the decision reached by the case worker although, inevitably, any review would be lodged in ignorance of the precise problem with the application.[54]

---

[49] Then Minister for State for Immigration and Home Office.
[50] http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/eu-justice-subcommittee/brexit-citizens-rights/oral/103890.html

[51] https://www.childrenslegalcentre.com/wp-content/uploads/2019/03/EUSS-briefing_Mar2019_FINAL.pdf
[52] https://www.gov.uk/guidance/eu-settlement-scheme-uk-tax-and-benefits-records-automated-check
[53] https://www.gov.uk/guidance/eu-settlement-scheme-uk-tax-and-benefits-records-automated-check
[54] https://www.gov.uk/guidance/eu-settlement-scheme-apply-for-an-administrative-review

The aims of the automated decision-making process

74. What were the government's aim when implementing a system which relies so heavily on an automated algorithm? The "*Processing Level Memorandum of Understanding*" between the Home Office and the DWP concerning the Settled Status scheme indicates that the automated decision-making process is intended, at least in part, to achieve a number of objectives[55] -

> - Reduce the Home Office reliance on paper documentation with a potential of 3.6m eligible applicants
> - Reduce caseworker processing time
> - Reduce fraud and error
> - Improve customer journey

75. This document also suggests that the government anticipated that there would be an "*influx*" of applications suggesting that it was also influenced by a need to create a system which could quickly address a large number of applications within a short period of time.

76. The document, "*EU Settlement Scheme private beta testing phase 2 report*" issued in January 2019 also suggests[56] that an aim of the algorithm was to allow applicants to –

> rely on government-held data to prove their UK residence automatically, minimising the evidential burden on them

77. The "*EU Settlement Scheme: public beta testing phase report*" from May 2019[57] suggests that the government has been successful in achieving some of the aims outlined in paragraph 74 above. It explains, for example, that:

    a. 75% of applications using the app were able to prove their identity in under 10 minutes.

---

[55]https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data /file/790668/Home_Office_-_DWP_API_EU_Exit_MoU.PDF. The equivalent document for the HMRC contains the same information and is available here: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/f ile/790661/Home_Office_-_HMRC_API_EU_Exit_MoU.PDF

[56] https://www.gov.uk/government/publications/eu-settlement-scheme-private-beta-2

[57]https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data /file/799413/EU_Settlement_Scheme_public_beta_testing_phase_report.pdf

b.   79% of applicants found the process "*very or fairly easy*".

c.   73% of applicants did not need to submit additional data to prove residence after the algorithmic review process.

78. Equally, Caroline Nokes MP explained to the House of Lords Select Committee on the European Union, Justice Sub-Committee that "*We are currently at a point where we are processing at a rate faster than intake*"[58] which again suggests that the government was achieving its aim of processing, very quickly, a large amount of applications.

79. There is a *possible* different perspective here. Government statistics reveal that post November 2018 (when the right to administrative review was introduced), 128 administrative review applications had been received and processed by 16 April 2019 with a further 46 being outstanding. The government explains that the review process revealed that 99 applicants had been assigned Pre-Settled Status rather than Settled Status.[59] Further, the government has explained that of these 99 cases[60] –

> The vast majority of these had originally provided additional evidence of their eligibility for settled status with their application for administrative review.

80. This suggests that 99 applicants had been deemed by the algorithms to have less than five years continuous residence (hence the need to provide additional evidence of their eligibility) and that this had been wrong. However, without knowing how many applications were lodged overall during the relevant period[61], we cannot realistically assess whether this calls into question the effectiveness of the scheme. Moreover, even if the algorithm did have an unacceptably high "*miss rate*", this *may* be mitigated to some degree by the administrative review process if it efficiently addresses applicants' concerns especially bearing in mind the plain importance of ensuring certainty quickly of an individual's immigration status.

---

[58] http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/eu-justice-subcommittee/brexit-citizens-rights/oral/103890.html

[59] https://www.gov.uk/government/publications/eu-settlement-scheme-public-beta-testing-phase-report
[60] https://www.gov.uk/government/publications/eu-settlement-scheme-public-beta-testing-phase-report
[61] We do not believe that the government has published the number of applications from November 2018 to 16 April 2019 so as to allow us to calculate the overall percentage of false negatives.

81. Caroline Nokes MP provided slightly different figures when she gave evidence to the House of Lords Select Committee on the European Union, Justice Sub-Committee in July 2019. She explained that up to 31 May 2019, there had been 253 cases of administrative review where the applicant was challenging a grant of pre-settled status and that 231 were overturned to be replaced with settled status.

82. Caroline Nokes MP also said that "*I regard it as impossible for somebody to be refused on the grounds of a protected characteristic*". However, this bold statement may not stand up to scrutiny. Indeed, as further explained below, there is evidence to suggest that certain groups which are protected under the Equality Act 2010 may be particularly at risk of being incorrectly rejected for Settled Status. We should note here that we have not seen an equality impact assessment of this issue. We have no reason to believe that one has been carried out pursuant to the Equality Duty in section 149 of the Equality Act 2010.

Equality implications of excluding data on Child Benefits and/or Child Tax Credits

83. The algorithm used within the Settled Status process does not look at DWP data concerning Child Benefits and/or Child Tax Credits. The Coram Children's Legal Centre has identified that this decision can impact negatively on women.[62] Limiting the databases interrogated by the algorithm in this way, the Coram Children's Legal Centre reasons, places women at a disadvantage because they are more likely to be in receipt of Child Benefit (it is payable only to the primary parent) and/or Child Tax Credits than men.

Analysing the claim

84. In our Opinion, this system leads to the application of one or more PCPs:

- The algorithm which assesses five-year residency by examining benefits data;

- The decision to apply that algorithms to limited number of benefits data sets but excluding Child Benefit and/or Child Tax Credits.

---

[62] https://www.childrenslegalcentre.com/wp-content/uploads/2019/03/EUSS-briefing_Mar2019_FINAL.pdf

85. The system may indeed be analysed as being two overlapping PCPs. Accordingly, it must be asked: ***Does the application of these (one or more) PCPS give rise to unlawful direct discrimination?***

86. Answering this question first requires a detailed analysis of the application of section 19 to these PCPs. This requires consideration of each of the sub-sub-paragraphs of section 19 (2) which we have set out above at paragraph 36.

87. **Section 19 (2) (a)**: First, we understand that these PCPs, singly or in combination, are applied to all applicants regardless of gender; accordingly, we consider that section 19 (2) (a) is satisfied.

88. **Section 19 (2) (b):** Secondly, we must consider whether these PCPs put women at a particular disadvantage in relation to men thereby satisfying section 19 (2) (b). There seems to be good reason to conclude that this may be the case. Thus we understand that in August 2018, 87% of Child Benefit recipients were female and 12% were male.[63] Equally, there is clear statistical evidence released in January 2018 showing that women are more likely to be in receipt of Child Tax Credits whether they be single, parents or in a couple.[64] The remaining question is whether this difficulty is readily cured or not by the paper trail demands.

89. Importantly, we note here that the process of providing documentary evidence is not necessarily straight forward and as such there is a meaningful disadvantage experienced by people who are required to provide such a paper trail. An extract from an article[65] which collated anecdotal evidence of such an applicant's experiences explains how this disadvantage occurs -

> People who were asked to bring in more evidence went into the standard panic mode which almost everyone experiences when dealing with the Home Office. "Many were unable to identify which piece of paper was the most relevant one," a Roma support group reported. "As a result, people came with piles of papers including council tax bills, GPs appointments, bank account statements… with no idea about which one to choose. The overall process was

---

[63]https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/782370/ChB_18_commentary_pdf.pdf

[64] See figure 7.1 at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/677582/cwtc-main-Dec17.pdf

[65] https://www.politics.co.uk/blogs/2019/02/06/warning-lights-flashing-over-eu-settled-status-app

> therefore extremely time consuming, with two hours on average spent per application while some taking up to four hours and over several sessions."
>
> The same was true of more tech-literate people contacting me on Twitter. The desire to send the Home Office everything to make absolutely sure the application was safe was very common.
>
> "I spent about two weeks gathering up all the documents I needed that proved five continuous years of residence - I was advised by a co-worker who had done five successful work visa applications to gather and submit everything I had. I even downloaded six years of bank statements, got together every single proof of entry into the country dating back to 2012, and tracked down every tenancy agreement I had over the years."
>
> This is the problem with the evidence workload. The Home Office is trying to limit the items people send to ten, but applicants will be nervous and keen to make absolutely sure their status is secure so they'll send in as much as they can, increasing the demands on caseworkers.

90. We consider that women would have been placed at a particular disadvantage and section 19 (2) (b) would be satisfied, if it were to be shown by relevant statistical evidence that women are at a greater risk of needing to demonstrate Settled Status via a paper trail. We emphasise that our view is contingent on this risk being actually established.

91. **Section 19 (2) (c):** Next it would be necessary to show that there is a victim, who has experienced the disadvantageous impact. We consider that a woman in receipt of Child Benefit and/or Child Tax Credits, who had been rejected by the automated algorithms and required to produce paper evidence despite having five years continuous residence, would satisfy section 19 (2) (c) of the Equality Act 2010. In theory it is enough to establish a real disadvantage, yet it is sensible also to consider the extent of the disadvantage. This will be significant later when proportionality comes to be considered. It is obvious that it will be much easier to justify a measure that causes little disadvantage as opposed to one that causes a good deal. Such disadvantage would be relatively modest if the woman had actually been able to meet the demand for a paper trail easily.

92. If a woman can properly assert, that these three conditions for a *prima facie* case of unlawful Prohibited Conduct of indirect discrimination on the basis of the Protected Characteristic of gender, contrary to section 19, are fulfilled, she could

commence a claim in the County Court pursuant to section 114 of the Equality Act 2010. It would then be up to the government to make its defence.

Analysing the possible defence

93. The government's defence could seek to challenge any or all of the points we have made above. This is likely to entail a forensic analysis of the evidence which we have considered as well as a legal argument about the nature of a PCP.

94. However, that is not the limit of the possible defence arguments that might be put forward. As we have noted above, section 19 (2) (d) provides for the possibility of a justification defence. It should be emphasised that it is for the defendant to assert and prove that justification and not for the claimant to assert and prove that there is no such justification. In short, the burden of proof would be on the government to justify its use of these PCPs, by establishing a legitimate aim for using them and that they are an effective and proportionate means of doing so.

95. We cannot completely analyse how the government might go about this for the simple reason that we have not seen any such pleading and there has not yet been an exchange of pre-action correspondence in which they might be fully set out.

96. Nonetheless, in our Opinion, we consider that the government would find it difficult to justify its use of these PCPs if it were to rely on the aims outlined at paragraph 74 above (i.e. reducing the Home Office's reliance on paper documentation, reducing caseworker processing time, reducing fraud and error and improving the customer journey). Whilst, as we explained at paragraphs 79 to 80 above, we cannot currently assess the effectiveness of the system overall in terms of accurately identifying who does have five years continuous residency, it does appear that any justification defence would fail since there are means of achieving the government's aims which are *less or non-discriminatory*. Put quite simply, it is difficult to understand why it did not interrogate the Child Tax Credit and/or Child Benefit database in the same way as the other benefits data base. This would, it appears, cure the discriminatory impact of the current system under discussion here whilst presumably improving its accuracy. Whilst it is recognised that one of the aims underpinning the PCPs is that a system was required which would quickly address a large number of applications within a short period of time, it is difficult to see why interrogating this extra data base via the algorithm would have caused any significant delay or other difficulty. Moreover, it is to be expected that utilising more relevant data under the control of the government

would lead to more accurate results and a lower need to require additional paper documentation. In other words, it is surely a step which would have improved rather than hindered the system which, if correct, would almost certainly be fatal to any justification defence.

97. We should acknowledge that there may be technical reasons for the decision not to interrogate this data which would have to be assessed by the court. We are aware that Parliamentary Questions were asked around this topic. On 6 June 2019, Paul Blomfield MP asked the following question[66] -

> To ask the Secretary of State for the Home Department, pursuant to the Answer of 20 May 2019 to Question 254186 and with reference to page 9 of the Memorandum of Understanding between his Department and the Department for Work and Pensions and to pages 12-13 of the Memorandum of Understanding between his Department and HMRC, for what reasons automated residency checks will not check applicants' records on working tax credit and child tax credit.

98. On 13 June 2019, Caroline Nokes MP indicated that the relevant data was not available for interrogation[67] -

> Her Majesty's Revenue and Customs advice indicated that receipt of working tax credit and child tax credit are not currently available as a data set that we can interrogate through the automated checks.

99. This prompted Paul Blomfield MP to ask a further question[68] in response on 20 June 2019 –

> To ask the Chancellor of the Exchequer, pursuant to the Answer of 13 June 2019 to Question 261211 on Immigration: EU Nationals, for what reason the information requested is not recorded in a data set that that can be interrogated through the automated checks for the EU Settlement Scheme.

---

[66] https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2019-06-06/261211/
[67] https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2019-06-06/261211/
[68] https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2019-06-20/267306/

100. In response, on 25 June 2019, Elizabeth Truss MP acknowledged[69] that such data was actually available but sought to suggest that it would only be relevant to a small number of people –

> HMRC and the Home Office have worked closely together to ensure that information provided for the automatic residence checks is sufficient to provide a response for the majority of applications. After taking account of the other information sources available to the Home Office, tax credit records have not been included as they would only be relevant to a small number of applicants. Individuals that received tax credits during a period of residence that is not validated automatically will still be able to provide documents directly to the Home Office as evidence of their residence.

101. On 26 June 2019, Paul Blomfield MP then asked[70] for details of how many applicants were accordingly in receipt of tax credits. He received the following surprising response on 3 July 2019 –

> The purpose of the automated checks is to help the applicant establish their continuous residence in the UK. Eligibility to receive benefits is not dependent upon being continuously resident in the UK. Benefits except Job Seekers Allowance and Maternity Allowance on their own are not strong indicators of continuous residence unless receipt of them persists over a period of 12 months. There are also overlaps with other data that is available from the checks. For instance, PAYE data covers most applicants who claim working tax credits and receipt of other benefits included in the checks will overlap with receipt of child tax credits. Child Benefit is not included in the automated checks because it is not a sufficient indicator of continuous UK residence. A full explanation of how the automated checks work has been published at
> https://www.gov.uk/guidance/eu-settlement-scheme-uk-tax-and-benefits-records-automated-check.
>
> Following analysis conducted on an anonymised sample of 10,000 applications submitted under the Immigration (European Economic Area) Regulations 2006, the Home Office estimated that the potential pool of resident EEA citizens who might benefit from tax credits data being included in the automated checks was around two per cent and this was before any consideration of the applicability of that data to proving continuous residence. Individuals who need to rely on tax credits to demonstrate their continuous residence are still able to provide documentary evidence of this as part of their application.

---

[69] https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2019-06-20/267306/
[70] https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2019-06-26/269726/

ai LAW

CLOISTERS

> In accordance with section 149 of the Equality Act 2010, we have had due regard to the Public Sector Equality Duty.

102. This is a surprising response because if 2% is representative of the number of applicants who receive tax credits, this would suggest that relevant data for at least 70,000 people[71], a significant number, was not interrogated, without any real justification.

103. On 16 July 2019, Lord Anderson raised this precise issue with Caroline Nokes MP when she gave evidence to the House of Lords Select Committee on the European Union, Justice Sub-Committee.[72] Her response confirms that a significant number of people receive tax credits and there appears to be no reason for excluding this relevant data -

> **Lord Anderson of Ipswich:** Minister, on the subject of tax credits, which I know you are familiar with through previous Answers you have given, I understand that automated residency checks do not check applicants' records on working tax credit, despite the fact that residency needs to be established as a condition for working tax credit. Liz Truss said on 25 June that, "tax credit records have not been included as they would only be relevant to a small number of applicants". You then gave an Answer on 3 July, saying that, after analysing 10,000 applications, "the Home Office estimated that the potential pool of resident EEA citizens who might benefit from tax credits data being included in the automated checks was around two per cent".
>
> On my calculations, 2% of 3 million is about 60,000 people, so, on the face of it, tax credit records could be usefully integrated into the scheme. Is there anything more you would like to tell us? First, do you really think that 60,000 is a small number, and, secondly, can you help us with the real reason why it was not possible to integrate tax credits? I am open to any number of reasons why it might have been.
>
> *Caroline Nokes MP:* I do not regard 60,000 as a small number at all, but, of those 60,000, many will also have HMRC records, and they might have DWP records as well. It was simply not one of the functionalities included. There is no hidden reason.

---

[71] It is estimated by the government that there was 3.5 million resident EU citizens: https://www.gov.uk/government/news/home-office-publishes-report-on-eu-settlement-scheme-public-beta-test-phase

[72] http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/eu-justice-subcommittee/brexit-citizens-rights/oral/103890.html

> **Lord Anderson of Ipswich:** But these are people who, according to your Answer, might benefit from the residence data. If you are going to tell me that there is some appalling technical obstacle to incorporating those data or that the data are not very reliable, I will understand. What I do not understand is, if it can easily be done, why for the sake of those 60,000 people it is not being done.
>
> *Caroline Nokes MP:* I am not going to say that I do not trust the data. It would be entirely inaccurate to portray it as that. It has not been included, and it is always something that I am prepared to consider including, but we have built the scheme now and it is working.
>
> **Lord Anderson of Ipswich:** Are you saying that it could not be retrofitted?
>
> *Caroline Nokes MP:* I do not know whether it can be retrofitted. That is something I can take away and look at, but I do not know if we can or not.
>
> **The Chairman:** Perhaps you could respond to us on that important question.

104.   In our Opinion, excluding this relevant data, for such a significant number of people for apparently no good reason, is unjustifiable in light of the disparate impact on women. That is, the aims of the Settled Status scheme are not being achieved in a proportionate way, because by broadening the databases interrogated by the algorithm in a simple and achievable way would have improved the accuracy of the algorithm, achieved more accurate results, and stopped or significantly stopped the disparate impact on women.

105.   We refer also to our discussion of the FAST principles at paragraph 61 above. It would appear that the principles of transparency and accountability have been ignored as revealed by the various exchanges above. This would present a further difficulty for the government in demonstrating the proportionality of its decision to use these PCPs.

Equality implications of considering only the main applicant for a benefit

106.   The government website is clear that whilst an applicant will not be informed of the reason that an automated check has concluded that the person does not have five years' continuous residence, one reason may be that[73] –

---

[73] https://www.gov.uk/guidance/eu-settlement-scheme-uk-tax-and-benefits-records-automated-check

> … you were not the main applicant on a benefit claim, for example you were a dependent on a Housing Benefit or Universal Credit claim

107. This suggests that the algorithm has been designed only to examine the main applicant for a benefit or only has access to data as to who is the main applicant. Accordingly, the Coram Children's Legal Centre reasons that this is a further way in which women are placed at a disadvantage by the Settled Status scheme since men are more likely to be the main applicant of Universal Credit which is one of the benefits examined.[74] At present we are unable to identify any statistics to that effect, but on the assumption that this is correct, we can readily see the equality implications of Coram's assertion if this is the case.

Analysing the claim

108. We consider that the decision to focus only on the main applicant would almost certainly be a PCP within section 19 (1) of the Equality Act 2010. It also appears that the PCP would be applied to men and women so that section 19(2) (a) is met. Next if, as a result, it made it harder, in a statistically significant way, for women to prove residence, then the PCP would place women at a particular disadvantage so that section 19 (2) (b) was also satisfied. It would then be necessary to identify one or more women who had been affected so that section 19(2) (c) was met. If so, then as for the previous example, there would be a *prima facie* case of Prohibited Conduct by way of indirect discrimination on the grounds of gender which was made unlawful by section 29 of the Equality Act 2010.

The defence

109. Subject to any forensic points that the government might take in relation to the PCP (as we have noted in the previous example), the key question for the defence would be whether the PCPs could be justified within the meaning of section 19(2) (d) of the Equality Act 2010. We repeat that the burden of establishing this defence would be on the government and that we cannot wholly anticipate all that it may wish to say on this account. However, we can comment on how it might be expected to proceed.

---

[74] https://www.childrenslegalcentre.com/wp-content/uploads/2019/03/EUSS-briefing_Mar2019_FINAL.pdf

110.   We currently have no information as to why the government would deem it appropriate or sensible to only focus on the main applicant for a benefit. Since Universal Credit examines household income, the data pertaining to a secondary applicant must be recorded somewhere.[75]   In the absence of that information, it is arguable that excluding relevant data i.e. the secondary applicant for Universal Credit, from the information examined by the algorithm is unjustifiable in light of any disparate impact on women. That is, the aims of the Settled Status scheme are not being achieved in a proportionate way, if, by simply broadening the databases interrogated by the algorithm would improve the accuracy of the algorithm, thereby better achieving the aims, whilst stopping any disparate impact on women. Equally, the breach of the FAST principles outlined at paragraph 61 above are likely to be highly relevant to proportionality.   The lack of transparency in this area would make it much harder for the government to mount a successful justification defence.

Disabled applicants and their carers

111.   The Coram Children's Legal Centre has also argued that the Settled Status scheme is problematic for disabled people arguing that[76] -

> Disabled people and their carers who rely on welfare benefits will need to provide additional proof of residence. This places an additional burden on these groups who may struggle to provide relevant documentation.

112.   We have not been able to identify the precise basis for arguing that contending that disabled people and their carers who rely on welfare benefits will need to provide additional proof of residence, but if this is correct, then it gives rise to a number of potential claims which we outline below -

   a. Disabled applicants may be able to bring an indirect disability discrimination claim if there is a PCP which is being applied here which gives rise to the need to provide additional information. It is possible that the government would be able to justify any *prima facie* indirect discrimination but without understanding more about the way in which disabled people are potentially disadvantaged by the system, we cannot comment any further.

[75] https://www.gov.uk/universal-credit/how-to-claim
[76] https://www.childrenslegalcentre.com/wp-content/uploads/2019/03/EUSS-briefing_Mar2019_FINAL.pdf

ai
LAW

CLOISTERS

b. Disabled applicants may also be able to bring a claim relying on section 15 of the Equality Act 2010 on the basis that they are being unfavourably treated by the demand for additional checks for a reason arising from their disability, namely being in receipt of welfare benefits. Again, the government would be entitled to seek to justify this unfavourable treatment. We cannot advise as to whether this would succeed without information as to the aims of the government in taking this approach and as to its success in meeting those aims.

c. Disabled applicants might be able to bring a claim on the basis that there has been a failure to make reasonable adjustments to the Settled Status scheme contrary to section 20 of the Equality Act 2010. That is, they are placed at a substantial disadvantage (by the additional checks) which should be alleviated through a modification to the system.

d. We add that is possible that the carers of a disabled applicant might be able to bring an "*association*" claim. In some circumstances, the Equality Act 2010 allows individuals to bring claims for direct discrimination if they are treated less favourably because of their *association* with a disabled person. An example of such a claim is *C-303/06 Coleman v Attridge Law (A Firm)*[77] where a woman was treated less favourably by her employer because her child was disabled in circumstances where she would have been treated better had her child been non-disabled. There is no basis upon which direct discrimination by association can be justified.

e. Alternatively, depending on the precise factual matrix, it may be that carers of disabled people could bring an indirect discrimination by association claim. The jurisprudence in relation to these types of claims is less well developed, but the main authority is *C-83/14 CHEZ Razpredelenie Bulgaria AD v Komisia za zashtita ot diskriminatsia*[78] in which the CJEU concluded that a non-Roma shopkeeper in a region of Bulgaria was subjected to discrimination because she lived in a predominantly Roma region and accordingly the government provided an inferior electricity service to all residents regardless of ethnicity and which disadvantaged her due to her

---

[77] EU:C:2008:415; [2008] E.C.R. I-5603; [2008] 3 C.M.L.R. 27; [2008] All E.R. (EC) 1105; [2008] C.E.C. 986; [2008] I.C.R. 1128; [2008] I.R.L.R. 722.
[78] EU:C:2015:480; [2016] 1 C.M.L.R. 14; [2015] All E.R. (EC) 1083; [2015] I.R.L.R. 746.

proximity and association with her Roma neighbours. This type of indirect discrimination by association claim would be subject to a justification defence.

113.   Should further information become available concerning the nature of the disadvantage which they suffer under Settled Status, we would be happy to explore further the extent to which disabled people and their carers might have Equality Act 2010 claims.

114.   We should add here that there is a limited exception with the Equality Act 2010 in relation to immigration status and disability contained within Schedule 2, Part 4. Specifically, government decisions taken which are "*necessary for the public good*" in relation to disability are not regulated. However, there is no reason to believe that this exception would apply to the present scenarios.

Other possible equality angles

115.   We also note that there may be other groups which are disparately impacted by the Settled Status process. For example, we understand that being in receipt of a pension bypasses the need to examine the five-year period,[79] thereby disadvantaging younger people. Caroline Nokes MP implied in July 2016 that students might also face difficulties satisfying the algorithm because they "*are not interacting with either the tax or the benefits systems*"[80] which again gives rise to a possible age angle. Moreover, commentators have identified that an individual under the age of 21 is unlikely to satisfy the five-year residency test more generally.[81] At the other end of life, it is possible that the requirement to engage with the entire Settled Status process online places older people at a disadvantage on the basis that they can be less computer literate[82] although it is noted that the use of "*pop up application centres*" may have mitigated this impact to a degree. There is also a suggestion that applicants who undertake informal seasonal work were at an increased risk of needing to provide additional documentation to verify their residential status and this may link to certain Protected Characteristics.[83]

---

[79] A link to the document entitled, "*Government records we use to confirm your UK residence*":
https://www.gov.uk/guidance/eu-settlement-scheme-uk-tax-and-benefits-records-automated-check
[80] http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/eu-justice-subcommittee/brexit-citizens-rights/oral/103890.html
[81] http://www.infiniteideasmachine.com/2019/01/Automated_data_checks_in_the_EU_Settlement_Scheme.pdf
[82] https://www.ageuk.org.uk/information-advice/work-learning/technology-internet/
[83] https://www.politics.co.uk/blogs/2019/02/06/warning-lights-flashing-over-eu-settled-status-app

However, we have not addressed these equality issues any further in this paper since they do not arise from the government's decision to deploy automated decision-making processes which is the focus of our instructions.

Risk-based verification (RBV)

116.   We turn now to the second case study identified by Swee Leng Harris, concerning Housing Benefits and Council Tax Benefits administered by LAs.   LAs are required under legislation to determine an individual's eligibility for these benefits. There is no fixed verification process but LAs can ask for documentation and information from any applicant "*as may reasonably be required*".[84] Since 2012, the DWP has allowed LAs to voluntarily adopt RBV systems as part of the application process as part of this verification process and has given guidance as to how this may happen.

117.   In its Housing Benefit and Council Tax Benefit Circular on Risk-Based Verification of HB/CTB Claims Guidance, [85] the DWP describes RBV as being –

> a method of applying different levels of checks to benefit claims according to the risk associated with those claims

118.   We understand that the way in which RBV works is to assign a risk rating to each applicant for Housing Benefit and Council Tax Benefit which determines the level of identity verification required. This allows the LA to target and focus resources on "… *those cases deemed to be at highest risk of involving fraud and/or error*".[86] For example, an individual with a low risk might simply need to provide proof of identity but someone with a high-risk rating might be subject to Credit Reference Agency checks, visits, increased documentation requirements etc.[87] It is very possible that the delays associated with these more stringent checks could

---

[84] The Council Tax Benefit Regulations 2006, reg 72:
http://www.legislation.gov.uk/uksi/2006/215/regulation/72/made and the Housing Benefit Regulations 2006, reg 86:: http://www.legislation.gov.uk/uksi/2006/213/regulation/86/made
[85] Housing Benefit and Council Tax Benefit Circular, HB/CTB S11/2011:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/633018/s11-2011.pdf
[86] Ibid.
[87] Ibid.

ai LAW

CLOISTERS

have dire consequences such as homelessness. Accordingly, it is important to be clear about precisely how individuals are assigned to different risk groups.

119. We know that LAs are free to choose algorithmic software to perform this task, and that ML is likely to be involved. The DWP circular[88] shows that the Department is aware that ML algorithms are being deployed -

> 12. Some IT tools use a propensity model which assesses against a number of components based on millions of claim assessments to classify the claim into one of the three categories above. Any IT system must also ensure that the risk profiles include 'blind cases' where a sample of low or medium risk cases are allocated to a higher risk group, thus requiring heightened verification. This is done in order to test and refine the software assumptions.
>
> 13. Once the category is identified, individual claims cannot be downgraded by the benefit processor to a lower risk group. They can however, exceptionally, be upgraded if the processor has reasons to think this appropriate.

120. However, it has been impossible to identify any publicly available information that explains *how* such algorithms are being deployed, or *on what basis*. This lack of transparency is a problem with all algorithm AI and ML systems that are not explained as we have noted already in relation to HART. We do not know, but it would certainly seem to be possible that any such algorithm applies discriminatory criteria as we further explain at paragraph 124 below.

121. The DWP Circular[89] states that LAs are required to "*audit*" their use of RBV systems and we are aware that at least some LAs are performing Equality Impact Assessments on their RBV systems. This is of course to be encouraged and is certainly good practice. Nonetheless, we have seen documentation that suggests that LAs are adopting a deeply unsophisticated approach. For instance, Central Bedfordshire Council's Audit Committee report of 9 April 2018 suggests that the LA should be satisfied that its ML algorithm was not discriminatory merely because the mathematical model did not look directly at any protected characteristic[90] -

---

[88] Ibid.
[89] Ibid.
[90] https://centralbeds.moderngov.co.uk/documents/s77223/08%20Annual%20Review%20of%20Risk%20Based%20Verification%20RBV%20Policy%20for%20Housing%20Benefit%20and%20Local%20Council%20Tax%20S.pdf

> **Equalities Implications**
>
> 40. Pursuant to the Equality Act 2010 ("the Act"), the Council, in the exercise of its functions, has to have due regard to (i) eliminating discrimination, harassment, victimisation and any other conduct that is prohibited by or under the Act; (ii) advancing equality of opportunity between those with a relevant protected characteristic and those without; and (iii) fostering good relations between those with a relevant protected characteristic and those without. The relevant protected characteristics are age, race, disability, gender reassignment, pregnancy and maternity, religion or belief, sex and sexual orientation. The duty also covers marriage and civil partnership, but to a limited extent.
>
> 41. RBV will apply to all New Claims for Housing Benefit and Localised Council Tax Support. A mathematical model is used to determine the Risk score for any claim. This model does not take into account any of the protected characteristics dealt with by the Equalities Act.
>
> 42. The Risk score profiles are determined by a propensity model; a mathematical formula which uses historical outcome data to establish the likelihood of fraud and error appearing in any given claim. Each benefit claim is analysed by the risk score software to identify if any of the characteristics associated with the occurrence of fraud and error are present. Likelihood is expressed by a risk category of high, medium or low risk'.
>
> 43. The course of action to be taken in respect of the risk score is governed by this policy. As such there should not be any equalities impact.

122. This is a very superficial assessment. It does not get under the skin of the process at all. It ignores the facts that the algorithm might be looking at proxies for protected characteristics (like postcode) or it may have "*learnt*" rules that are indirectly discriminatory. It shows a failure to understand how equality law and the use of algorithms, AI and ML can interact. We have not conducted an audit of local authorities to look for similar reasoning, but this Council's approach does raise questions as to whether it is a typical or unique failure to get to grips with the issues we have identified above.

123. Another example of poor practice occurred in 2015, when Rochdale Borough Council, misunderstood the difference between indirect and direct discrimination, concluding that its RBV system did not give rise to an equality issues simply because "*it will be applied equally to all claims*".[91] The report added that "*the use of*

---

[91]http://democracy.rochdale.gov.uk/documents/s38489/Append.%202%20for%20Risk%20Based%20Verification%20Policy.pdf

RBV software solution to risk score each claim effectively eliminates any officer prejudices in deciding what level of scrutiny is used in assessing each claim". The Council has erroneously assumed that AI is free from, or necessarily eliminates, discrimination.

124.   We must add that we do not think that it is fanciful to consider that RBV will be giving rise to discrimination in some instances. For instance, the Bedfordshire Audit Committee report noted that the ML algorithm it used for its RBV system consistently detected a far greater percentage of "*high risk*" applicants than had been anticipated[92] -

Year	Detection vs Expectation
2015/2016	33% vs 20%
2016/2017	33% vs 20%
2017/2018	33% vs 20%

125.   When a random sample of 10 of the "*high risk*" applicants were further examined, they were all found to be women who were working. This could be a coincidence, as the sample was small, or it could suggest that the algorithm had "*learnt*" a discriminatory correlation. It ought to have rung alarm bells, since it is well-established from studies of AI that pattern recognition technology can unintentionally lead to the replication of human biases in various subtle ways. For instance, the House of Commons Science and Technology Select Committee noted this in 2018, pointing out how ML algorithms can, far from introducing objectivity, actually perpetrate discrimination[93] -

> **Bias**
>
> 32. While sharing data widely is likely to improve the quality of the algorithms they support, the underpinning systems also need to produce reliable and fair results— without bias. Machine learning is "application agnostic". Algorithms are designed to discriminate—to tell the difference—between, for example, people, images or documents. As Professor Louise Amoore of Durham University explained, "in order for an algorithm to operate, it has to give weight to some pieces of information over others", and this bias is "intrinsic to the algorithm". Durham Constabulary warned against demanding "some hypothetical perfection", and instead suggested considering "the conditions that would persist if such models were not available". Dr Pavel Klimov, Chair

---

[92]Ibid.

[93] "*Algorithms in decision-making*", House of Commons Science and Technology Committee Fourth Report of Session 2017–19 Report, 15 May 2018, HC 351: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf

ai LAW

CLOISTERS

of the Law Society's Technology and the Law Group, highlighted the importance of not turning the technology into "a weapon against ourselves", referring to the need for checks and balances. Some forms of bias can nevertheless extend beyond what is acceptable. Although algorithms have the potential to "promote efficiency, consistency, and fairness", they can also "reinforce historical discrimination or obscure undesirable behaviour".

33. The Alan Turing Institute told us that when automated decision-making is applied "current legislation does very little to protect individuals from being discriminated" against. Where algorithms are used in the criminal justice system it is imperative that algorithms are not unfairly discriminatory. This is not always the case. We were told by the Information Commissioner that "algorithmic risk scores used in some US states" to determine sentencing "inaccurately classified black defendants as future criminals at almost twice the rate as white defendants, perpetuating a bias that already existed in the training data."   Even in relatively benign uses of algorithms such as when advertisements are displayed online, the result can be that "users of that service are being profiled in a way that perpetuates discrimination, for example on the basis of race".

34. Oxford and Nottingham Universities warned that as the complexity of algorithmic applications increases, "so do the inherent risks of bias, as there is a greater number of stages in the process where errors can occur and accumulate". If discrimination (of the undesirable type) is introduced, subsequent deployment can amplify the discriminatory effects. Discrimination can enter the decision-making process from a variety of paths— in the use of inappropriate 'training data', a lack of data, through correlation disguised as causation, or the unrepresentativeness of algorithm development teams—and can present itself at any stage of an algorithm's lifecycle including conception, design, testing, deployment, sale, or its repurpose.

Training data

35. Perhaps the biggest source of unfair bias is inappropriate 'training data'— the data from which the algorithm learns and identifies patterns and the statistical rules which the algorithm applies. The way that training data are selected by algorithm developers can be susceptible to subconscious cultural biases, especially where population diversity is omitted from the data. The Royal Society noted that "biases arising from social structures can be embedded in datasets at the point of collection, meaning that data can reflect these biases in society". A well-recognised example of this risk is where algorithms are used for recruitment. As Mark Gardiner put it, if historical recruitment data are fed into a company's algorithm, the company will "continue hiring in that manner, as it will assume that male candidates are better equipped. The bias is then built and reinforced with each decision." This

is equivalent, Hetan Shah from the Royal Statistical Society noted, to telling the algorithm: "Here are all my best people right now, and can you get me more of those?" Microsoft told us that, as part of its 'Fairness, Accountability and Transparency in Machine Learning' initiative, computer scientists were examining how some recruitment algorithms had learned biases "based on a skewed input data". During our inquiry, Professor Louise Amoore of Durham University informed us of the case of a black researcher at MIT working with facial-recognition algorithms who found that "the most widely used algorithms did not recognise her black face". Professor Amoore explained that the AI had been trained to identify the patterns in a facial geometry using predominantly white faces.

36. Professor Nick Jennings from the Royal Academy of Engineering believed that algorithms are "not always well trained because people do not always understand exactly how they work or what is involved in training". Because the research in this area is still relatively undeveloped, he explained, "you end up with poorly trained algorithms giving biased results". This vulnerability can be difficult to tackle when, as is increasingly the case, the process of compiling training data and the process of pattern-learning are separate endeavours. Machine learning algorithm developers can procure training data from third parties, such as data brokers, where "access to the original basis on which the data was collected is unavailable". The Horizon Digital Economy Research Institute explained that "as algorithms become embedded in off-the-shelf software packages and cloud services, where the algorithm itself is reused in various contexts and trained on different data, there is no one point at which the code and data are viewed together".

Insufficient data

37. As well as unrepresentative data, insufficient data can also cause discrimination. As prediction accuracy is generally linked to the amount of data available for algorithm training, incorrect assessments could be more common when algorithms are applied to groups under-represented in the training data.

…

Correlation without causation

38. Bias or unfairness can arise, the Royal Society told us, when a machine learning algorithm correctly finds attributes of individuals that predict outcomes, but "in contexts where society may deem use of such an attribute inappropriate". The Institute of Mathematics and its Applications gave the example of an algorithm used by the courts in Broward County, Florida, which asks: 'Was one of your parents ever sent to jail or prison?' Even if predictive,

the Institute emphasised the unfairness of the inference that "a defendant deserves a harsher sentence because his father went to prison".

39. The sophistication of pattern-learning means that even setting restrictions on the algorithms produced, for example to ignore protected characteristics like race, may not easily solve the problem. Machine learning systems may instead identify proxies for such characteristics. Professor Amoore explained how in the US, where algorithms had been used to predict the outcome in criminal trials, "even where race as a category was removed from the input data, the algorithm still learned characteristics, or attributes, that we might say are in breach of the Equality Act, because they use what we could call proxies for race. They learn patterns in past patterns of crime or they learn patterns in postcodes, for example." Following a review of Durham Constabulary's HART algorithm, used to aid custody decisions (paragraph 21), a postcode field was removed amid concerns that it could discriminate against people from poorer areas. Concerns have been expressed that other characteristics used in HART and other policing algorithms are potential sources of bias, especially where they serve as proxies for race or gender. (paragraph 41).

40. The opaque nature of the algorithm 'black box' makes its use controversial in some areas. Professor Amoore warned that there may exist "areas of our social, political or economic lives where we might want to say there is no place for algorithmic decision making". She also questioned the use of inference and correlation in the criminal justice system, and suggested that its use in the US for sentencing "constitutes a violation of due process or overt discrimination". (In the UK, Durham Constabulary was using an algorithm to help determine whether a low-risk offender is suitable for 'deferred prosecution'.)

(footnotes removed)

126.   We consider that every LA that applies RBV and every Audit Committee and any relevant Auditors that review its use, should be made fully aware of the learning points in the passage above; the House of Commons Committee there explains why and how algorithms can unintentionally discriminate.   As Jennifer Cobbe eloquently puts it in her paper[94] -

"Overcoming the assumption that decisions made by machines must be rational, whilst a psychological step rather than a legal one, is important.   Unless reviewers accept that ADM [Automated Decision Making] systems can produce irrational results, no assessment of whether an ADM system has in fact produced an irrational result can take place.   In reviewing ADM systems, it will therefore be important to hold them to the same standards as humans, lest imperfect

---

[94] Ibid.

> systems by permitted to make potentially problematic decisions without the
> appropriate scrutiny".

The problem of the "*black box*"

127.   The inability to peer inside an algorithm, AI or ML process, so as to understand
how decisions are being made, is sometimes called the "*black box*" problem. It has
real legal implications. In equality law it is not enough for an employer or service
provider to say that we don't believe that there is discrimination, but it is not *known*
what exactly has happened. The law expects transparency, so that individuals are
able to know with certainty whether or not discrimination is occurring. This is, in
reality, an aspect of the rule of law. The positive right to non-discrimination must
be capable of being ascertained by the court. In that sense the principle of equality
before the law goes hand in hand with the idea that transparency is an essential
ethical value that those using AI should adopt.

128.   In this section we shall explain in more detail what the transparency problem
is by considering it in the context of the RBV and we consider it further in the
context of public law principles later in this paper. By contrast, with the Settled
Status issues which we have already discussed above where publicly available
information provides some basis for concluding that discrimination could be
occurring,[95] in relation to RBV, we can only speculate that discrimination *might* be
at play. In those circumstances, how can an individual know if discrimination is
occurring and equally how can an LA be sure that it is not?

Opening the "*black box*" to identify the existence or absence of discrimination?

129.   We shall start by looking at this question in practical terms from the point of
view of a person claiming and entitled to Housing Benefit and Council Tax
Benefits.

130.   We consider that are perhaps four ways in which transparency might occur
although they are not mutually exclusive.

---

[95] Because we know that the automated decision-making algorithm is only looking at certain benefits
databases and statistically speaking this will impact on women.

Voluntary disclosure

131.   It is possible, though it seems very unlikely, that a LA might voluntarily explains how an algorithm works to an applicant. It seems very unlikely because we suspect that most LAs will not understand the algorithm that is used since it will have been provided by a third party. For instance, in the Bedfordshire Audit Committee report referred to above, it was explained that the LA was largely dependent on information provided by the third party for its understanding of what was happening -

> We have been advised by our RBV software supplier that the variation in risk scores is quite normal as the original distribution percentage was calculated in 2011 when RBV was introduced nationally and due to changes in caseload demographics that have taken place over the last four years, specifically the increase in the number of in-work claims, there has been an increase in the number of high risk cases. Our RBV software supplier has analysed the outturn from our new claims and whilst our number of high risk cases are higher than the expected average, the results show we receive a proportionally higher number of claims from claimants who do not receive social security benefits. There is a greater propensity for such claims to be deemed as "higher risk". A very small number of claims from pension age applicants were found to be high risk.

132.   This could be a fair assessment, or it could be a self-serving piece of advice to justify an apparent anomaly between outturn and expectation that was in fact the result of a discriminatory design fault or a discriminatory input.

Outcome based suspicion

133.   Rather than voluntary disclosure, what seems more likely is that individuals or NGOs such as advice agencies, start to perceive a pattern as to whom is being subjected to additional checks. This may lead to a suspicion of discrimination. For example, it might be suddenly noticed that members of certain protected groups were being subject to higher scrutiny.   Each of us is aware that such a suspicion has justified the commencement of litigation, which in turn through the process of disclosure and analysis, has led to proof of unlawful discrimination.

134.   Thirdly, an organisation, perhaps with the benefit of crowdfunding or the backing of a litigation funder, might dedicate money and resources to identifying discrimination. This happened in the US when journalists at ProPublica analysed 7,000 "*risk scores*" to identify that a machine learning tool deployed in some states was nearly twice as likely to falsely predict that black defendants would be criminals in the future in comparison to white defendants.[96]

General Data Protection Regulation

135.   Lastly, individuals *might* identify discrimination through a request for transparency pursuant to article 15 of the General Data Protection Regulation (GDPR) in so far as personal data is being processed.[97]   This obliges a Data Controller to explain the categories of personal data being processed and in particular –

> the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject

136.   Some care should be taken in relying on this provision though. The former "*Article 29 Data Protection Working Party*", which was set up to give guidance on key guidance on aspects of the GDPR, authoritatively[98] concluded that the GDPR does ***not*** go so far as to dictate that algorithms or the basis for machine learning must be disclosed.[99]   Likewise, the guidance on automated decision given by UK's

---

[96] https://www.nytimes.com/2016/08/01/opinion/make-algorithms-accountable.html

[97] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679

[98] The Article 29 Working Party ceased to exist on 25 May 2018. It has been replaced by the European Data Protection Board: see https://edpb.europa.eu/. The EDPB has adopted this guidance: https://edpb.europa.eu/node/89.

[99] See "*Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*":
 https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

Information Commissioner's Office (ICO) suggests that the principle of transparency is fairly weak when it comes to algorithms, saying that[100] -

> You must inform individuals if you are using their data for solely automated decision-making processes with legal or similarly significant effects. This applies whether you have received the data directly from the individuals concerned or from another source.
>
> You must also provide meaningful information about the logic involved and what the likely consequences are for individuals. This type of processing can be invisible to individuals so in circumstances where it can have a significant impact on them you need to make sure they understand what's involved, why you use these methods and the likely results.
>
> How can we explain complicated processes in a way that people will understand?
>
> Providing 'meaningful information about the logic' and 'the significance and envisaged consequences' of a process doesn't mean you have to confuse people with over-complex explanations of algorithms. You should focus on describing: the type of information you collect or use in creating the profile or making the automated decision; why this information is relevant; and what the likely impact is going to be/how it's likely to affect them.
>
> **Example:**
>
> An on-line retailer uses automated processes to decide whether or not to offer credit terms for purchases. These processes use information about previous purchase history with the same retailer and information held by the credit reference agencies, to provide a credit score for an online buyer. The retailer explains that the buyer's past behaviour and account transaction history indicates the most appropriate payment mechanism for the individual and the retailer. Depending upon the score customers may be offered credit terms or have to pay upfront for their purchases.

137. We also agree with Swee Leng Harris that the existing Article 35 GDPR requirement to analyse processing that assesses the risks to the rights and freedoms of data subjects by means of a Data Protection Impact Assessment is of

---

[100] https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling-1-1.pdf

little utility because the only obligation is to provide it to the ICO and not to publish it generally.[101]

138.   In summary, no effective mechanism has yet been provided through the GDPR and the ICO to enable individuals to readily understand how decisions about them are being taken, as a result of this increasing use of automated governmental decision-making. This is not however the end of the matter.

The "black box" and the burden of proof in equality claims

139.   In equality law it is well established that a lack of transparency in a pay system can give rise to an *inference* of discrimination. This was established some thirty years ago in C-109/88 *Danfoss*[102] and has been reiterated on many occasions. So, paradoxically, the lack of meaningful transparency as to the way in which an algorithm or AI or ML works, might assist claimants who are prepared to litigate without having certain evidence of discrimination.   In our view, this principle of transparency is as applicable to the provision of goods, facilities and services, including public functions, as it is to employment context.

140.   If a claimant could persuade a court to draw an inference of discrimination from a lack of transparency, this would shift the burden of proof onto the defendant to prove that there is no discrimination.   Following European law, this reversal of the burden of proof has now been systematised into a general rule contained in section 136 (2) of the Equality Act 2010 which now provides[103] that -

> If there are facts from which the court could decide, in the absence of any other explanation, that a person (A) contravened the provision concerned, the court must hold that the contravention occurred.

141.   Accordingly, if a service provider is not able or willing to explain how an algorithm is operating, this might be sufficient to shift the burden of proof in a

---

[101] Harris, Swee Leng, Data Protection Impact Assessments as Rule of Law Governance Mechanisms June 3, 2019:
https://zenodo.org/record/3237865#.XTGSTPJKhQI
[102] Case C- 109/88, *Handels- og Kontorfunktionaerernes Forbund i Danmark v Dansk Arbejdsgiverforening Ex p. Danfoss A/S* EU:C:1989:383;[1989] E.C.R. 3199; [1991] 1 C.M.L.R. 8; [1991] I.C.R. 74; [1989] I.R.L.R.
[103] See *Wong v Igen Ltd (formerly Leeds Careers Guidance)* [2005] EWCA Civ 142; [2005] 3 All E.R. 812; [2005] I.C.R. 931; [2005] I.R.L.R. 258 for a discussion of the right way to apply the reversal of the burden of proof. See also *Efobi v Royal Mail Group Ltd* [2019] EWCA Civ 18, [2019] 2 All E.R. 917, [2019] I.C.R. 750; [2019] I.R.L.R. 352.

discrimination case from the claimant to the defendant to provide a non-discriminatory explanation.

142.   However, it is important to note, that we are not suggesting that a lack of transparency will shift the burden of proof in every situation.   There would have to be some reason to suspect at least, at some plausible level, that there may be discrimination. In a direct discrimination case that would mean that there would have to be at least proof of a Protected Characteristic and of disadvantageous treatment compared with persons not having that Protected Characteristic. There would then have to have been some attempt to secure transparency as to the reason for the treatment which had been rebuffed.

143.   In the case of indirect discrimination, the principle of transparency is likely to work in a different way. It will be remembered that at the justification stage the burden of proving a valid basis upon which to justify *prima facie* indirect discrimination rests squarely on the defendant. In the absence of transparency this will, as a matter of common sense, be extremely difficult. If a service provider cannot or will not explain how an algorithm, AI or ML works, in a particular context, there will likely be a great difficulty in establishing the key components of justification. If it is not explainable, how can it be shown to be effective? It might equally be asked how can it be shown to be proportionate?   It is not impossible that these questions might be answered positively by the service provider without being transparent as to what data base has been used and how the automated-decision making, AI or ML, works, but it is obvious that there will be significant difficulties. The answers offered would have to invite the court or tribunal to take a good deal on trust. In a case where there was little disadvantage shown this may be easier than one where there is proof of a substantial disadvantage to groups protected under the Equality Act 2010.

Justification and RBV

144.   Having considered the particular issue surrounding transparency, AI, ML and discrimination, we now turn to consider how a LA might justify RBV. Again, we repeat that we cannot anticipate every argument that might be deployed at this stage and there is a distinct lack of information concerning RBV which makes assessing any justification very difficult.

145.   The stated justification for RBV is that it allows LAs to identify fraudulent claims accurately, quickly and in a cost-efficient way.[104]   However, there is evidence to suggest that, at least in the past, some LAs that have utilised RBV do not consider it to be particularly effective in achieving these aims. We note that in 2012, only 53% of LAs thought that the RBV process had been effective in reducing fraud and error. The same low percentage thought that the RBV process had been effective in improving processing times.[105]   Accordingly, unless an LA can demonstrate that its RBV scheme is actually achieving its intended aims of greater accuracy and speed, it is possible that the justification defence will fail. Equally, in the absence of transparency, there has been a breach of the FAST principles, which as we explained at paragraph 61 above, might mean that the proportionality element of the justification test cannot be satisfied.

GDPR consequences to discriminatory automated decision making

146.   Finally, we should also explain here that under Article 22 of the GDPR there is a right *not* to be subject to fully automated decision making, in certain circumstances, where it produces legal effects concerning the individual, or a similarly significant affect.

Article 22

Automated individual decision-making, including profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

2. Paragraph 1 shall not apply if the decision:

(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

(b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

(c) is based on the data subject's explicit consent.

---

[104] Housing Benefit and Council Tax Benefit Circular, HB/CTB S11/2011; ibid.
[105] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/211088/847summ.pdf

> 3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.
>
> 4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

147.   This is important because it appears that at least some LAs might be implementing RBV which is fully automated, that is, with no human involvement at all. What is more, it appears that some LAs are implementing processes that occur without an individual being aware of it and maybe without their consent. This is an extract from Waltham Forest's "*Equality Analysis – Risk Based Verification policy*" dated 29 May 2015:[106]

> The Council's RBV software is integrated with the benefit processing software and intelligent online electronic forms. When a customer submits an online claim / application or change of circumstances form, it is immediately risk scored via an automated link and an email is returned confirming the level and type of evidence required.
>
> **The entire RBV process is fully automated and seamless, it occurs within the processing software so does not rely on manual intervention by benefits officers. It also takes place without the knowledge of the person submitting the form.**
>
> (Emphasis added)

148.   Does Article 22 prevent the use of RBV Machine Learning algorithms? This will depend on whether a process of seeking additional identification in order to access a benefit is sufficiently significant to fall within Article 22 (1). The ICO's guidance suggests if the algorithm was being used to determine *eligibility* for a welfare benefit then Article 22 (1) would apply, but that simply seeking additional documentation probably does not:[107]

---

[106]https://democracy.walthamforest.gov.uk/documents/s47042/Equality%20Analysis%20RBV%20Policy%20V1%20APPENDIX%202.pdf

[107] https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling-1-1.pdf

> **What types of decision have a legal or similarly significant effect?**
>
> A decision producing a legal effect is something that affects a person's legal status or their legal rights. For example when a person, in view of their profile, is entitled to a particular social benefit conferred by law, such as housing benefit.
>
> A decision that has a similarly significant effect is something that has an equivalent impact on an individual's circumstances, behaviour or choices.

149.   However, if it can be demonstrated that the RBV process gives rise to discrimination contrary to the Equality Act 2010, then the necessary legal effect would be created and Article 22 (1) would apply. This would mean that the RBV system was unlawful.   This is also confirmed by the ICO guidance.[108] Accordingly, it is important to remember that any Equality Act 2010 claim arising from fully automated algorithms, should also be accompanied by a claim for breach of Article 22.

150.   There is a final point to make here which links to the extent to which any RBV system can be justified under section 19 of the Equality Act 2010. As we explained at paragraph 49 above, in order for any justification defence to succeed, it must be underpinned by a legitimate aim and certainly means an aim which is in itself lawful.   In our view, if RBV is being undertaken in a way which is *prima facie* discriminatory and if it is fully automated contrary to Article 22, then there can be no legitimate aim and any justification defence must necessarily fail.

C) Public Sector Equality duty and automated decision making

Outline

151.   The two examples we have been asked to review in this Opinion, as a basis for our more general advice on the effect of algorithms AI and ML based decision-making, concern the exercise of public power. That means that these decisions must conform to public law rules and will be subject to review if they do not. Whilst a consideration of these public law principles is outside the remit of this opinion, it is important to consider here the Public Sector Equality Duty, which is

---

[108] https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling-1-1.pdf

often analysed alongside public law principles. Some of these public law issues are examined in the excellent paper by Jennifer Cobbe entitled "*Administrative law and the machines of government: judicial review of automated public-sector decision-making*"[109].

152.   All public authorities are subject to section 149 of Equality Act 2010 which created the new consolidated Public Sector Equality Duty (PSED) applying to all the Protected Characteristics in the Equality Act 2010 –

> (1) A public authority must, in the exercise of its functions, have due regard to the need to—
>
> (a) eliminate discrimination, harassment, victimisation and any other conduct that is prohibited by or under this Act;
>
> (b) advance equality of opportunity between persons who share a relevant protected characteristic and persons who do not share it;
>
> (c) foster good relations between persons who share a relevant protected characteristic and persons who do not share it.
>
> (2) A person who is not a public authority but who exercises public functions must, in the exercise of those functions, have due regard to the matters mentioned in subsection (1).
>
> (3) Having due regard to the need to advance equality of opportunity between persons who share a relevant protected characteristic and persons who do not share it involves having due regard, in particular, to the need to—
>
> (a) remove or minimise disadvantages suffered by persons who share a relevant protected characteristic that are connected to that characteristic;
>
> (b) take steps to meet the needs of persons who share a relevant protected characteristic that are different from the needs of persons who do not share it;
>
> (c) encourage persons who share a relevant protected characteristic to participate in public life or in any other activity in which participation by such persons is disproportionately low.
>
> …

---

[110] [2014] Eq LR 60, [2013] EWCA Civ 1345.

> (5) Having due regard to the need to foster good relations between persons who share a relevant protected characteristic and persons who do not share it involves having due regard, in particular, to the need to—
>
> (a) tackle prejudice, and
>
> (b) promote understanding.
>
> (6) Compliance with the duties in this section may involve treating some persons more favourably than others; but that is not to be taken as permitting conduct that would otherwise be prohibited by or under this Act.
>
> (7) The relevant protected characteristics are—…race …
>
> …
>
> (9) Schedule 18 (exceptions) has effect.

153.   The PSED imposes a duty to advance equality of opportunity, and so imposes clear obligations on public authorities such as ministers and government departments to understand who the persons might be who may be affected by their actions "*who share a relevant protected characteristic*" and suffer "*disadvantages… connected to that characteristic*", and remove or minimise those disadvantages: section 149(3)(a); also to identify "*the needs of persons who share a relevant protected characteristic that are different from the needs of persons who do not share it*", and "*take steps to meet*" those needs: section 149(3)(b). The PSED is therefore concerned with the outputs from the exercise of public functions. This is why it is now commonplace for public authorities to carry out equality impact assessments even though these are not specifically required by the law. More generally a failure to comply with the PSED fully can render a decision unlawful. There is extensive jurisprudence on the obligations that arise from the PSED which were usefully summarised by McCombe LJ in *Stuart Bracking & Ors v Secretary of State for Work and Pensions*[110] –

> 25…I summarise the points [in the jurisprudence].
>
> (1) As stated by Arden LJ in *R (Elias) v Secretary of State for Defence* [2006] 1 WLR 3213; [2006] EWCA Civ 1293 at [274], equality duties are an integral and

---

[110] [2014] Eq LR 60, [2013] EWCA Civ 1345.

important part of the mechanisms for ensuring the fulfilment of the aims of anti-discrimination legislation.

(2) An important evidential element in the demonstration of the discharge of the duty is the recording of the steps taken by the decision maker in seeking to meet the statutory requirements: *R (BAPIO Action Ltd) v Secretary of State for the Home Department* [2007] EWHC 199 (QB) (Stanley Burnton J (as he then was)).

(3) The relevant duty is upon the Minister or other decision maker personally. What matters is what he or she took into account and what he or she knew. Thus, the Minister or decision maker cannot be taken to know what his or her officials know or what may have been in the minds of officials in proffering their advice: *R (National Association of Health Stores) v Department of Health* [2005] EWCA Civ 154 at [26 – 27] per Sedley LJ.

(4) A Minister must assess the risk and extent of any adverse impact and the ways in which such risk may be eliminated before the adoption of a proposed policy and not merely as a "rearguard action", following a concluded decision: per Moses LJ, sitting as a Judge of the Administrative Court, in *Kaur & Shah v LB Ealing* [2008] EWHC 2062 (Admin) at [23 – 24].

(5) These and other points were reviewed by Aikens LJ, giving the judgment of the Divisional Court, in *R (Brown) v Secretary of State for Work and Pensions* [2008] EWHC 3158 (Admin), as follows:

i) The public authority decision maker must be aware of the duty to have "due regard" to the relevant matters;

ii) The duty must be fulfilled before and at the time when a particular policy is being considered;

iii) The duty must be "exercised in substance, with rigour, and with an open mind". It is not a question of "ticking boxes"; while there is no duty to make express reference to the regard paid to the relevant duty, reference to it and to the relevant criteria reduces the scope for argument;

iv) The duty is non-delegable; and

v) Is a continuing one.

vi) It is good practice for a decision maker to keep records demonstrating consideration of the duty.

(6) "[G]eneral regard to issues of equality is not the same as having specific regard, by way of conscious approach to the statutory criteria." (per Davis J (as he then was) in *R (Meany) v Harlow DC* [2009] EWHC 559 (Admin) at [84],

approved in this court in *R (Bailey) v Brent LBC* [2011] EWCA Civ 1586 at [74–75].)

(7) Officials reporting to or advising Ministers/other public authority decision makers, on matters material to the discharge of the duty, must not merely tell the Minister/decision maker what he/she wants to hear but they have to be "rigorous in both enquiring and reporting to them": *R (Domb) v Hammersmith & Fulham LBC* [2009] EWCA Civ 941 at [79] per Sedley LJ.

(8) Finally, and with respect, it is I think, helpful to recall passages from the judgment of my Lord, Elias LJ, in *R (Hurley & Moore) v Secretary of State for Business, Innovation and Skills* [2012] EWHC 201 (Admin) (Divisional Court) as follows:

(i) At paragraphs [77–78]

"[77] Contrary to a submission advanced by Ms Mountfield, I do not accept that this means that it is for the court to determine whether appropriate weight has been given to the duty. Provided the court is satisfied that there has been a rigorous consideration of the duty, so that there is a proper appreciation of the potential impact of the decision on equality objectives and the desirability of promoting them, then as Dyson LJ in *Baker* (para [34]) made clear, it is for the decision maker to decide how much weight should be given to the various factors informing the decision.

[78] The concept of 'due regard' requires the court to ensure that there has been a proper and conscientious focus on the statutory criteria, but if that is done, the court cannot interfere with the decision simply because it would have given greater weight to the equality implications of the decision than did the decision maker. In short, the decision maker must be clear precisely what the equality implications are when he puts them in the balance, and he must recognise the desirability of achieving them, but ultimately it is for him to decide what weight they should be given in the light of all relevant factors. If Ms Mountfield's submissions on this point were correct, it would allow unelected judges to review on substantive merits grounds almost all aspects of public decision making."

(ii) At paragraphs [89–90]

"[89] It is also alleged that the PSED in this case involves a duty of inquiry. The submission is that the combination of the principles in Secretary of State for *Education and Science v Tameside Metropolitan Borough Council* [1977] AC 1014 and the duty of due regard under the statute requires public authorities to be properly informed before taking a decision. If the relevant material is not available, there will be a duty to acquire it and this will frequently mean than

> some further consultation with appropriate groups is required. Ms Mountfield referred to the following passage from the judgment of Aikens LJ in Brown (para [85]):
>
> '….the public authority concerned will, in our view, have to have due regard to the need to take steps to gather relevant information in order that it can properly take steps to take into account disabled persons' disabilities in the context of the particular function under consideration.'
>
> [90] I respectfully agree…."

154. We think it is obvious from this that if there is no assessment of the discriminatory impact that may arise from the use of automated decisions, AI and ML there may well be a failure to comply with section 149.

Interplay between AI, transparency and the Public Sector Equality Duty

155. This leads to an important point: what if no meaningful assessment of the discriminatory impact of automated decisions, AI and ML occurs because of the opacity problem which we identified at the outset of this paper in paragraph 12 above?

156. This appears to have been one of the difficulties encountered in the recent case of *R v The Chief Constable of South Wales Police ex parte Bridges* [2019] EWHC 2341 (*Bridges*). In that case, it was recognised that South Wales Police were required under section 149 of the Equality Act 2010 to have regard to "*the need to eliminate discrimination harassment, victimisation …*" (paragraph 150). In 2017, South Wales Police did undertake an initial assessment of whether deploying facial recognition technology could lead to direct discrimination but apparently it omitted to examine whether indirect discrimination could occur (paragraphs 151 – 152). Mr Bridges argued that in those circumstances there was a breach of section 149 of the Equality Act 2010. The court was critical of this submission, noting that it had "*an air of unreality*" because "*there is no firm evidence that the software does produce results that suggest indirect discrimination*" (paragraph 152). However, what the court did not grapple with sufficiently is that the lack of "*firm evidence*" was due to a lack of transparency within the system deployed by the police. This point was made by Dr Jain, an expert witness in the case, who is quoted in the judgment at paragraph 155 as saying -

> I cannot comment on whether AFR Locate has a discriminatory impact as I do not have access to the data sets on which the system is trained and therefore cannot analyse the biases in those data sets.
>
> …
>
> Before it is possible to draw conclusions on the existence of gender bias, an extensive study needs to be conducted where match scores are thoroughly analysed for both male and females regardless of whether they generate alerts or not. Mr Edgell [witness called by the police] does not carry out that study; he considers only alert statistics.

157. There are two points here. First, there is the question as to the database on which the system was trained and secondly the possibly of checking it with a "*study*". As to the first, it is not clear why there could not have been full disclosure of the system so that the original training databases could have been examined. There are proprietary systems which are said to be able to detect bias in training data bases such as IBM's AI Fairness 360 Open Source Toolkit.[111] IBM described this Toolkit by saying –

> This extensible open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. Containing over 70 fairness metrics and 10 state-of-the-art bias mitigation algorithms developed by the research community, it is designed to translate algorithmic research from the lab into the actual practice of domains as wide-ranging as finance, human capital management, healthcare, and education. We invite you to use it and improve it.

158. For the future it is perhaps worth noting that in early 2019, IBM produced a data set of facial images taken from a Flickr dataset with a 100 million photos and videos in order to improve the accuracy, and hopefully remove bias from, facial recognition technology.[112]

159. As to the second issue, presumably, the practical problem identified by Dr Jain has been caused, if not exacerbated by, the fact that data is immediately deleted if it does not create an alert. This brings the opacity problem that we identified above at paragraphs 12 and 13 into sharp relief. AI and ML is frequently characterised

---

[111]

http://aif360.mybluemix.net/?utm_campaign=the_algorithm.unpaid.engagement&utm_source=hs_email&utm_medium=email&utm_content=69523284&_hsenc=p2ANqtz-9vaujms_IQeQkh8nE92xGK7pisSc5eYX3nQkytSKQkCd7rAAd2pPmn_kgregFKWVMMD7G0LuVo_jhLB1G1fQZNL81PKA&_hsmi=69523284

[112] https://www.cnbc.com/2019/01/29/ibm-releases-diverse-dataset-to-fight-facial-recognition-bias.html

by a lack of transparency through design or because of fears about misusing data. But, it is this opacity which allows discrimination to fester. It renders the Public Sector Equality Duty largely toothless if the courts permit opacity to essentially defeat an argument that the discriminatory impact of technology has not been adequately assessed. It is precisely where there is opacity that a discriminatory impact assessment is most needed. The court does not appear to have been recognised that the lack of transparency itself might be indicative of discrimination in a way which is analogous to *Danfoss*, a case that we discuss at paragraph 139 above.

160. There is a further problem with the analysis in *Bridges*. That is, the court appears to conclude that the facial recognition technology deployed by the police is unlikely to be discriminatory because of the "*important failsafe*" of a police officer verifying the match proposed by the automated facial recognition technology (paragraph 156). However, this ignores the problem of automation bias, a phenomenon that we explain at paragraph 18 above by which human actors are more likely to accept the conclusion of a machine. It is not fanciful to be concerned that this happens.

161. Indeed, a report from July 2017 by The Independent Chief Inspector of Borders and Immigration reveals that an automated decision-making system which allowed applications to be "*streamed*" by risk may well have been susceptible to this form of automation bias[113] -

> 3.6 Since 2015, UKVI has been developing and rolling out a 'streaming tool' that assesses the perceived risks attached to an application. The streaming tool is regularly updated with data of known immigration abuses (for example, data relating to breaches of visa conditions after entry to the UK). It streams applications 'Green' (low risk), 'Amber' (medium risk) or 'Red' (high risk). In Istanbul, a fourth rating, 'Super Green', had been in use. This was described as "Applications with evidenced low risk, where limited or no judgement was required and that could be streamed to AO6 grade decision makers.
>
> 3.7 While segmenting applications in order to manage them more efficiently is sensible given the numbers involved, there is a risk that the streaming tool becomes a de facto decision-making tool. This is a particular concern for applications streamed 'Green', where the benchmark is for 70-75 visit visa decisions per day per ECO, and where the quality assurance regime is geared

---

[113]https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/631520/An-inspection-of-entry-clearance-processing-operations-in-Croydon-and-Istanbul1.pdf

to checking only anomalous decisions (for example the refusal of a 'Green' or issue of a 'Red'). The assurance regime does not take account of the danger of 'confirmation bias' (an unconscious disinclination on the part of the decision maker to look for or give appropriate weight to evidence that contradicts the streaming rating, and a tendency to select and rely on evidence that supports it).

3.8 At the Croydon DMC, in the first two months of 2017, less than 4% of 'Green' applications were refused. Meanwhile, nearly 50% of all visit applications streamed as 'Red' were issued, plus over 80% of those streamed as 'Amber'. The 'Green refusals' and 'Red issues' add up to over 3,600 decisions that 'deviated' from the streaming rating. With only five ECMs in post, managers will have struggled with this volume of decisions to quality assure, and this may go some way to explaining why inspectors found that other management tasks had not been completed.

162. In short, we take the view that when it comes to AI / ML and section 149 of the Equality Act 2010, a "*rigorous consideration of the duty, so there is a proper appreciation of the decision on equality objectives*" as endorsed in *Bracking* (above) can probably only occur where there is sufficient transparency[114] so that bias can be monitored and that the inclusion of a human actor should not be sufficient to allow an organisation to side-step this obligation due to the problem if automation bias. Accordingly, we consider that *Bridges* requires further judicial scrutiny so as not to set a dangerous precedent.

D) Legal avenues

163. Parts 1 and 2 of Chapter 9 of the Equality Act 2010 concern bringing civil claims for individual remedies. Claims concerning goods facilities and services allegedly unlawful contrary to section 29 of the Equality Act 2010 are heard in the County Court, so a claim by an individual, for example, that the Settled Status process breached the Equality Act 2010 would ordinarily be litigated in that way. Any compensation would be limited to injury to feelings, perhaps around £3,000, unless the claimant could demonstrate a financial loss arising from an incorrect decision. Any evidenced financial losses which directly flowed from the discriminatory decision could also be recouped via a compensation award. The time limits for

---

[114] Whilst we recognise the importance of organisations processing data with great care, we consider that a carefully crafted system whereby data was retained for a short period of time at appropriate intervals so as to "*test*" for equality issues would be compliant with the GDPR.

bringing such claims is ordinarily six months though there are also provisions for secondary time limits: see section 118 of the Equality Act 2010.

164.   Such litigation is relatively rare largely because the relationship between costs and potential quantum is so disproportionate. Legal aid is almost always unavailable and there is little opportunity to have significant protection against costs. The Equality and Human Rights Commission can and sometimes does assist in bringing test cases. Recently the House of Commons Women and Equalities Committee has encouraged it to use these powers more.[115]

165.   It may also be possible to seek a judicial review. Chapter 9 of the Equality Act 2010 expressly reserves this possibility while conferring exclusive jurisdiction on the County Court for individual claims.[116] However, detailed consideration of this option is outside the remit of this opinion.

166.   Lastly, a complaint to the ICO in relation to any alleged failures to comply with the GDPR should not be ignored.

E)   Current debate and drawing together the issues

167.   AI and ML undoubtedly has enormous potential to further the public good. A recent illustration of the positive power of AI and ML has been highlighted by MIT[117] focusing on Danial Tso, a researcher at Google, who had developed an algorithm that beat a number of trained radiologists in testing for lung cancer. Tse and colleagues trained a deep-learning algorithm to detect malignant lung nodules in more than 42,000 CT scans. The resulting algorithms turned up 11% fewer false positives and 5% fewer false negatives than their human counterparts.[118]

---

[115] See "Enforcing the Equality Act: the law and the role of the Equality and Human Rights Commission", House of Commons Women and Equalities Committee, Tenth Report of Session 2017–19, HC 1470, 30 July 2019; see
ttps://publications.parliament.uk/pa/cm201719/cmselect/cmwomeq/1470/1470.pdf

[116] See section 113(3) of the Equality Act 2010.

[117] https://www.technologyreview.com/f/613560/google-shows-how-ai-might-detect-lung-cancer-faster-and-more-reliably/
[118] "*End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography*" by Diego Ardila, Atilla P. Kiraly, Sujeeth Bharadwaj, Bokyung Choi, Joshua J.

168. Importantly, despite this reality, there has been very little litigation about how to use and regulate AI. Most significantly, facial recognition technology utilised by South Wales Police has been challenged by way of a judicial review supported by Liberty in *Bridges*. Beyond the Liberty judicial review, we are not aware of any other litigation in which AI is being challenged under equality and/or administrative law principles within the UK. We are currently reviewing the situation across Europe for Equinet and we shall publish any further information that we have as to challenges of this type on our website www.ai-lawhub.com in due course.

169. This lack of litigation is against a backdrop in which there has been little meaningful public debate concerning AI and equality principles:

    a. The equality bodies in the UK, namely the Equality and Human Rights Commission[119] and the Northern Ireland Equality Commission[120], have yet to publicly engage with this issue and we are not aware that it is part of any proposed work streams.

    b. The ICO has produced numerous papers and blogs concerning AI and has identified its potential to discriminate but we are not aware of any detailed reference to how the Equality Act 2010 actually regulates this area.[121]

    c. Similarly, whilst The Office for Artificial Intelligence (OAI) and the Government Digital Service (GDS), in partnership with The Alan Turing Institute, have analysed AI through the framework of "*ethics*" (which includes freedom from "*biases*" and "*discrimination*") [122], there has been no

---

Reicher, Lily Peng, Daniel Tse, Mozziyar Etemadi, Wenxing Ye, Greg Corrado, David P. Naidich & Shravya Shetty is available here: https://www.nature.com/articles/s41591-019-0447-x

[119] https://www.equalityhumanrights.com/en

[120] https://www.equalityni.org/Home

[121] See, for example, "*Big data, artificial intelligence, machine learning and data protection*" available at https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf and the blogs published in 2018-2019 in advance of the ICO's proposed auditing framework for Artificial Intelligence at https://ai-auditingframework.blogspot.com/2019/03/simon-mcdougall-director-for-technology.html

[122] See, in particular, the government guidance "Understanding artificial intelligence ethics and safety" which is available at https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety and The Alan Turing Institute's "*Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*" available at https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf.

guidance from the government concerning how the Equality Act 2010 regulates AI in any detailed way.

d. In the academic sphere, it is common to read brief or fleeting references to the importance of non-discrimination and perhaps the Equality Act 2010[123], but there has been very little by way of detailed analysis of how the Equality Act 2010 actually applies to algorithmic decision making. The only detailed analysis that we are aware of, beyond our own work as outlined at www.ai-lawhub.com relates to the broader European context rather than being UK specific, for example, the work of Professor Sandra Wachter[124] of The Alan Turing Institute.

e. Whilst journalists[125] have identified the issues around AI in the public sector, it has yet to feature regularly in the mainstream news.

170.   In our view, a debate is required which focuses on the following issues, all of which we have identified above: legitimate aims, transparency, GDPR compliance (including the ICO's role), the lack of apt and comprehensive statutory guidance, and the role the Equality and Human Rights Commission in securing enforcement of equality law

171.   This debate is urgently required because there has been a relatively recent explosion in the use of automatic decision-making across the globe.[126] Indeed, the UK government is intent on embedding AI within the public sector and has committed within its Industry Strategy to funding AI initiatives in order to "*provide the government with innovative solutions for more efficient public service*"[127]. Recent

---

[123] For example, this paper from RUSI - https://rusi.org/sites/default/files/20180329_rusi_newsbrief_vol.38_no.2_babuta_web.pdf

[124] Her profile is available here: https://www.turing.ac.uk/people/researchers/sandra-wachter. One important paper from Professor Wachter in which she analyses European discrimination law is contained in her paper from June 2019: "*Affinity Profiling and Discrimination by Association in Online Behavioural Advertising*" available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3388639

[125] For example, see: https://news.sky.com/story/sky-views-algorithmic-austerity-cant-be-ended-by-money-11765341

[126] The Artificial Intelligence Index Report 18 is available here: http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf. It deploys a wide range of metrics to calibrate the extent to which AI is deployed globally.

[127] "*Industry Strategy: Artificial Intelligence Sector Deal*", updated May 2019, which is available here: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal__4_.pdf

research conducted by Sky News and Cardiff University identified that 53 LAs are using algorithms to predict behaviour including Bristol City Council which uses its "*Bristol Integrated Analytics Hub*" to analyse data relating to benefits, school attendance, crime, homelessness, teenage pregnancy and mental health from 54,000 local families to predict which children could suffer from domestic violence, sexual abuse or go missing.[128]   This research also identified that Kent Police now only investigates 40% of crime, as opposed to 75%, on the basis of predictive algorithms.   As the Special Rapporteur on extreme poverty and human rights, Philip Alston, explained in his report on the UK dated 12 April 2019:[129]

> Benefit claims are made online and the claimant interacts with authorities primarily through an online portal. The British welfare state is gradually disappearing behind a webpage and an algorithm, with significant implications for those living in poverty.

172.   Moreover, commentators are suggesting that governmental reliance on algorithms will only increase as a means of delivering services against a backdrop of austerity.[130]   It appears that the two case studies which we analyse above are simply the tip of the iceberg.[131]

173.   Accordingly, we entirely endorse TLEF's view that the interplay between automated data processing in decision making, particularly through AI and Machine Learning, is an area where there are gaps in legal analysis and there is a need for increased understanding of the application of the law.

ROBIN ALLEN QC

ra@cloisters.com

DEE MASTERS

deemasters@cloisters.com

---

[128] https://news.sky.com/story/the-controversial-tech-used-to-detect-problems-before-they-happen-11649080

[129] https://undocs.org/en/A/HRC/41/39/Add.1

[130] https://news.sky.com/story/sky-views-algorithmic-austerity-cant-be-ended-by-money-11765341

[131] By way of example, the government lists eight case studies of the way in which it is using artificial intelligence here: https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector#examples-of-artificial-intelligence-use

www.cloisters.com

www.ai-lawhub.com

7 September 2019