House of Commons Science and Technology Committee

Governance of AI – Call for evidence

Public Law Project's Response

Introduction

1. Public Law Project (PLP) is an independent national legal charity founded in 1990 with the aim of improving access to public law remedies for marginalised individuals. Our vision is a world in which the state acts fairly and lawfully. Our mission is to improve public decision-making, empower people to understand and apply public law, and increase access to justice. We deliver our mission through five programmes: litigation, research, advocacy, communications, and training. One of PLP's five strategic priorities for 2022-25 is ensuring that Government use of new technologies is transparent and fair. Scrutinizing the use of automated decision-making (ADM) systems and big data by Government is a key part of our work. Given our specific expertise, this evidence focuses on the use of ADM systems by public, rather than private, bodies.
2. We are not opposed in principle to Government use of ADM systems and we recognise their potential benefits. But given that it is a rapidly expanding practice, and of increasing importance in the lives of our beneficiaries, we are focused on ensuring that such systems operate fairly, lawfully and in a non-discriminatory way.
3. In 2022, PLP ran two roundtables on regulating Government use of AI, with over twenty participants, including civil society organisations, grassroots organisations, academics and individuals affected by ADM tools. Our response draws on the key themes arising from the roundtable discussions, as well as our research.

Executive summary

4. PLP is concerned about the impact of ADM on the lives of our beneficiaries and is focused on ensuring that such systems operate transparently, reliably, lawfully, fairly, in a non-discriminatory way, and with adequate authorisation in law for their use. Public trust in ADM systems with wide societal impacts requires that these tools are shown to work, and that they do so in accordance with the law and in respect of non-discrimination and other individual rights.
5. The existing legal framework governing the use of AI provides a patchwork of vital provisions that require ADM systems to be transparent and not to discriminate and/or breach other individual rights. However, we consider the governance of AI, which is necessary to ensure the that the legal framework works in practice, to be operate less effectively. Current governance of AI in the UK does not provide an adequate level of transparency, accountability, or protection against unlawfulness, unfairness, and discrimination.
6. We recognise the potential of the 'Algorithmic Transparency Standard' (ATS) to increase transparency of the public sector's use of algorithms. However, this potential will remain limited if engagement with the ATS is not made compulsory and if improvements are not made to the detail of information required to be published to allow individuals to properly understand the decision-making process to which they are subjected (paras 7-11).
7. Although fragmented, provisions under the Freedom of Information Act 2000, Data Protection Act 2018, UK GDPR, Equality Act 2010, Human Rights Act 1998 and ECHR provide a patchwork of vital, albeit imperfect, safeguards. If the legal framework were to be reformed, the focus should be on fortifying existing safeguards, ensuring clarity and coherence between existing laws and securing quick and effective ways of enforcing individual rights (paras 12-14).
8. In our view, a compulsory algorithmic transparency regime should be overseen by an independent regulator, such as the ICO or, a dedicated AI regulatory body. The regulation

of AI should be based on the key principles of: anti-discrimination, reflexivity, respect for privacy and data rights, meaningful transparency, accountability, and avenues for redress. Quick and effective avenues for redress for affected individuals are essential and could be achieved through a specialist, yet accessible regulator and a forum for complaints relating to Government (paras 15-18).

9. Positive examples of AI governance can be found within the EU Commission's AI Act proposal, Canada's Directive on Automated Decision Making (DADM), and France's Law for a Digital Republic. In our view, it is important that the UK draws inspiration from the compulsory transparency regimes found in Canada and France, under which individuals are notified when automation is used to reach a decision, and enough information is provided for individuals to understand the decision-making process to which they are subjected.
10. If the Government's White Paper on AI is to reform the legal framework, the focus should be on fortifying existing safeguards and ensuring clarity and coherence between existing laws. In our view, the law already requires transparency in Government use of automation, but much more needs to be done to secure meaningful transparency from public bodies operating tools that can have significant social impact. In this regard, inspiration could be drawn from other jurisdictions with compulsory transparency regimes, such as Canada and France. Furthermore, it is essential that there are quick and effective ways of enforcing existing rights within any reform to the governance of AI.

How effective is current governance of AI in the UK?

11. PLP has found that Government uses ADM in a wide range of high impact areas, including immigration, welfare benefits, policing and prisons, education, and more. To date, PLP has gathered more than forty examples of Government ADM systems through our investigative research. We will be publishing full details of these systems in our forthcoming 'Tracking Automated Government' register.

12. Government use of ADM technologies can mean quicker and more consistent decision-making. However, it also comes with significant risks and drawbacks. In our view, current governance of AI in the UK is ineffective to the extent that is has not succeeded in securing an adequate level of transparency; accountability; and protection against unlawfulness and unfairness, including ensuring privacy and data rights and protection against discrimination.

    a. **Lack of transparency** – To have trust in algorithms, particularly those with a wide societal impact, we need transparency and explainability: a trustworthy algorithm should be able to "show its working" to those who want to understand how it came to its conclusions.[1] We are observing however that in the UK Government use of ADM systems is marked by a high degree of opacity. For example, PLP is concerned about the use by the Home Office's Marriage Referral and Assessment Unit (MRAU) of an automated 'triage tool' to decide whether to investigate potential 'sham' marriages. For convenience, we will refer to this as the 'sham marriages algorithm'. The triage tool uses 8 risk factors unknown by the individuals subject to its processing.[2]

    PLP is also concerned by the Department for Work and Pensions' (DWP) lack of transparency around their use of automation. Their 2021-2022 accounts revealed that they are trialling a new machine learning-driven predictive model to detect possible fraud in Universal Credit claims. This model has already been used in relation to advances claims already in payment and the DWP expects to trial the model on claims before any payment has been made early in 2022-23. However, we are concerned that the DWP has not published any Equality Impact Assessments, Data Protection Impact Assessments or other evaluations

[1] See further D. Spiegelhalter, 'Should We Trust Algorithms', Harvard Data Science Review, Issue 2.1, Winter 2020. This paper was referenced by the Centre for Data Ethics and Innovation in its report 'Review into bias in algorithmic decision-making', November 2020.

[2] Three of the eight criteria have been disclosed in an Equality Impact Assessment received by PLP in response to a Freedom of Information request, however we consider that this disclosure is not sufficient in helping individuals understand how the algorithm is operating.

completed in relation to its automated tools. Given the lack of transparency, and especially the DWP's failure to publish equality analyses of any of its automated tools, we consider that the roll-out of their machine-learning model is premature.

These are just two examples of a broader trend. PLP's experience is that when asked to disclose further information on the development and operation of ADM tools through requests under the Freedom of Information Act 2000 (FOIA), both the Home Office and the DWP often refuse disclosure. Such refusals purport reliance on exemptions under FOIA, most often section 31(1)(a), which exempts information which, if released, would likely prejudice the prevention and detection of crime (we expand on this further below under 'To what extent is the legal framework for the use of AI fit for purpose'). Specifically, both departments rely on the vaguely formulated argument that disclosure of further information would permit individuals to 'game the system'. However, to our knowledge, the risk of gaming has not been further substantiated or particularised. Literature on the subject suggests that some features are not capable of being gamed, for example, when the features used by the algorithm are fixed and unalterable. Examples are race, sex, disability or nationality. Commentators note that disclosure of criteria not based on user behaviour offer: 'no mechanism for gaming from individuals who have no direct control over those attributes'.[3] Furthermore, disclosure of the criteria alone is generally insufficient for a user to 'game' the algorithm. Authors Cofone and Strandburg state:[4]

> Because effective gaming requires fairly extensive information about the features employed and how they are combined by the algorithm, it will often be possible for decision makers to disclose significant information about what

[3] N Diakopoulos, 'Accountability in algorithmic decision making' Communications of The Acm, February 2016, Vol. 59, No. 2

[4] Cofone, Ignacio and Strandburg, Katherine J, Strategic Games and Algorithmic Secrecy (2019) 64:4 McGill LJ 623

features are used and how they are combined without creating a significant gaming threat.

We are concerned that neither the Home Office nor the DWP have engaged with whether the information requested is in fact gameable and instead tend to give a blanket refusal to disclose on the basis of the risk of 'gaming'. We also emphasise that section 31 is subject to the public interest test, meaning that not only does the information have to prejudice one of the purposes listed but, before the information can be withheld, the public interest in preventing that prejudice must outweigh the public interest in disclosure. Our concern is that Government departments are not sufficiently engaging with a consideration of the public interest in transparency and disclosure, which we expand on further in this submission.

Opacity around the existence, details and deployment of ADM systems is a major challenge of working in this area is. There is opacity not only around the detail and impacts of such systems but also, in some cases, around their very existence. Doubtless, there are many other instances of Government automation, beyond the examples PLP has collected. But finding out about them is very difficult. The fruitfulness of PLP's investigative research has been largely dependent upon the willingness of Government departments to engage meaningfully with requests for information. We have found this to be patchy. Further, engagement with the Cabinet Office's Algorithmic Transparency Standard thus far appears to have been limited, with only six reports published to date despite the fact that there are many more than six tools currently in use in Government (see further below – 'What measures could make the use of AI more transparent and explainable to the public?').

Opacity is a cost in and of itself, undermining public trust in the use of ADM tools. It also comes with costs in terms of ensuring appropriate scrutiny and evaluation of such systems, including individuals being able to enforce their rights when

decisions made using them are unlawful. We consider that transparency is a prerequisite for accountability (see further below).

b. **Lack of accountability** – Transparency is a necessary starting point for evaluating new AI technologies and for accountability and redress.[5] Most importantly, individuals, groups, and communities whose lives are impacted should know they have been subjected to automated decision-making, especially given the costs, financially, emotionally, and physically, of a flawed AI system on the life of an individual.

   As it currently stands, affected individuals are rarely aware of the ADM system they have been processed by, making it much more difficult for them to obtain redress for an unfair or unlawful decision. Currently, PLP is particularly concerned about the 'sham marriages algorithm' in this regard. An Equality Impact Assessment disclosed by the Home Office in response to one of PLP's FOIA requests suggests that the algorithm appears to affect some nationalities more than others (see further below – 'Risk of discrimination'). But affected individuals rarely know about its existence.

   Public bodies ought not to be immunised from review, evaluation, scrutiny and accountability for their decisions. It cannot be the case that public bodies are capable of becoming so immunised in practice through adopting ADM systems that are so opaque as to make it practically impossible to assess how they are operating, whether they are doing so lawfully and fairly, and (if not) what may be done about it.

c. **Privacy and data protection costs** – Government ADM systems generally involve the processing of data on a large scale. This comes with widely recognised costs when it comes to privacy and data protection. In October 2019, the UN Special

[5] See Justice and Home Affairs Committee, 'Technology rules? The advent of new technologies in the justice system' (30 March 2022), available at: https://committees.parliament.uk/publications/9453/documents/163029/default/.

Rapporteur on extreme poverty and human rights noted that the digitisation of welfare systems poses a serious threat to human rights, raising significant issues of privacy and data protection.[6] In February 2020, a Dutch court ruled that a welfare fraud detection system, known as SyRI, violated article 8 (right to respect for private life, family life, home and correspondence) of the European Convention on Human Rights.[7] In September 2021, the UN High Commissioner for Human Rights produced a report devoted to privacy issues arising because of the widespread use of artificial intelligence. The report sets out problems including: the incentivization of collection of personal data, including in intimate spaces; the risk of data breaches exposing sensitive information; and inferences and predictions about individual behaviour, interfering with autonomy and potentially impacting on other rights such as freedom of thought and expression.[8]

In this regard, large scale data-matching exercises conducted by the UK Government, such as the DWP's General Data Matching Service[9] and the Cabinet Office-led National Fraud Initiative,[10] are of particular concern.

d. **Risk of discrimination** – Bias can be 'baked in' to ADM systems for various reasons, including as a result of problems in the design or training data. If the training data is unrepresentative then the algorithm may systematically produce worse outcomes when applied to a particular group. If the training data is tainted by historical injustices then it may systematically reproduce those injustices.

---

[6] UNCHR, 'Report of the Special rapporteur on extreme poverty and human rights' (11 October 2019) UN Doc A/74/493, available at https://undocs.org/A/74/493.
[7] *NJCM v The Netherlands* C-09-550982-HA ZA 18-388.
[8] UNCHR, 'Report of the United Nations High Commissioner for Human Rights: The right to privacy in the digital age' (13 September 2021), UN Doc A/HRC/48/31 available at https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session48/Documents/A_HRC_48_31_AdvanceEditedVersion.docx.
[9] Information about the General Data Matching Service can be found, for example, in the DWP's 'Fraud Investigations Staff Guide - Part 1', pages 135-6, available at https://www.gov.uk/government/publications/fraud-investigations-staff-guide.
[10] Information about the National Fraud Initiative is available at https://www.gov.uk/government/collections/national-fraud-initiative.

The possibility of problems with the training data was highlighted in the well-known *Bridges*[11] litigation, concerning the South Wales Police's (SWP) use of facial recognition technology. Before the High Court, there was evidence that, due to imbalances in the representation of different groups in the training data, such technologies can be less accurate when it comes to recognising the faces of people of colour and women.[12] The appeal was allowed on three grounds, one of which was that the SWP had not "done all that they reasonably could to fulfil the PSED [public sector equality duty]": a non-delegable duty requiring public authorities to actively avoid indirect discrimination on racial or gender grounds.[13]

There can also be bias in the design of an algorithm's rules. Following the initiation of a legal challenge by the Joint Council for the Welfare of Immigrants (JCWI) and Foxglove, the Home Office had to suspend a visa streaming tool used to automate the risk profiling it undertook of all applications for entry clearance, which used 'suspect nationality' as a factor in red-rating applications. Red-rated visa applications received more intense scrutiny, were approached with more scepticism, took longer to determine, and were more likely to be refused than green or amber-rated applications. JCWI and Foxglove argued that this amounted to racial discrimination and was a breach of the Equality Act 2010.[14]

Many of the ADM systems PLP is aware of appear to have an unequal impact on marginalised groups and/or groups with protected characteristics. For example, the Equality Impact Assessment disclosed for the sham marriages algorithm, referred to above, includes a graph showing that couples of Bulgarian, Greek, Romanian, and Albanian nationality are flagged for investigation at a rate of between 20% and

[11] *R (Bridges) v Chief Constable of South Wales Police and others* [2020] EWCA Civ 1058.
[12] See the expert report of Dr Anil Jain, available at https://www.libertyhumanrights.org.uk/wp-content/uploads/2020/02/First-Expert-Report-from-Dr-Anil-Jain.pdf.
[13] *R(Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058, [201].
[14] See 'We won! Home Office to stop using racist visa algorithm', JCWI Latest News blog, available at https://www.jcwi.org.uk/news/we-won-home-office-to-stop-using-racist-visa-algorithm.

25%. This is higher than the rate for any other nationality. By contrast, around 10% of couples involving someone of Indian nationality and around 15% of couples involving someone of Pakistani nationality fail triage and are thus, subject to a sham marriage investigation. No other nationalities are labelled on the graph.

To give another example, it appears that the DWP's automated tool(s) for detecting possible fraud and error in Universal Credit claims may have a disproportionate impact on people of certain nationalities. The Work Rights Centre have told us that, since August 2021, they have been contacted by 39 service users who reported having their Universal Credit payments suspended. Even though the charity advises a range of migrant communities, including Romanian, Ukrainian, Polish, and Spanish speakers, as many as 35 of the service users who reported having their payments suspended were Bulgarian, with three Polish and one Romanian-Ukrainian dual national.

e. **Other risks of unlawfulness and unfairness –** In addition to the risks of discrimination and data protection/privacy violations identified above, there are a number of common problems with ADM systems giving rise to other risks of unlawful and/or unfair decision-making. These problems include:

   i. Errors in the system's outputs. ADM systems can sometimes produce 'false positives': indications that something exists or is true when, in fact, it does not exist or is not true. Conversely, ADM systems can also produce 'false negatives': indications that something does not exist or is not true when, in fact, it does exist or is true. An ADM system will rarely, if ever, be 100% accurate and trade-offs often have to be made between reducing the number of false positives and reducing the number of false negatives. Put differently, design choices have to be made as to whether a system should be under-inclusive or over-inclusive. For example, the user of a fraud detection system might prefer that the system flags *all* but not *only* individuals who have actually committed

fraud, if the alternative is a system that flags *only* but not *all* such individuals i.e., they may prefer an over-inclusive system which produces false positives. This design choice may mean that people who have not committed fraud are nonetheless subjected to stressful fraud investigations. The potential for ADM systems to be over-inclusive and produce a high number of false positives is illustrated by the Australian 'Robodebt' scheme, under which hundreds of thousands of people were issued with computer-generated debt notices, some of which made demands for payment from people who did not actually owe the Australian Government any money.[15] A class action against the scheme resulted in over $1.7 bn in financial benefits being due to approximately 430,000 individuals.[16]

ii. Inflexibility arises from the fact that ADM systems often work by applying fixed rules uniformly across a broad range of cases. Unlike a human decision-maker, an ADM system cannot make an exception for a novel case that was not envisioned or foreseen when the system was developed. It can only act in accordance with its programming.

iii. Automation bias: a well-established psychological phenomenon whereby people put too much trust in computers.[17] This may mean that officials over-rely on ADM systems that were designed to *support* not *replace* human decision-making (sometimes known as 'decision support systems'), and fail to exercise meaningful review of the system's outputs. For example, the Independent Chief Inspector of Borders and Immigration found potential

[15] See for example Jordan Hayne and Matthew Doran, 'Government to pay back $721m as it scraps Robodebt for Centrelink welfare recipients' (29 May 2020) ABC News, available at https://www.abc.net.au/news/2020-05-29/federal-government-refund-robodebt-scheme-repay-debts/12299410. The Australian Government established the Royal Commission into the Robodebt Scheme on 18 August 2022 and their final report is due in April 2023. Information about the Royal Commission is available at https://robodebt.royalcommission.gov.au/.

[16] See: https://gordonlegal.com.au/robodebt-class-action/

[17] See, for example, L.J. Skitka and others, 'Does automation bias decision-making?'(1999) 51 International Journal of Human-Computer Studies 991.

evidence of automation bias in the use of the aforementioned visa streaming tool. At the Croydon immigration Decision Making Centre (DMC) in the first two months of 2017, less than 4% of 'green' rated applications were refused. Meanwhile, nearly 50% of all visit applications rated 'red' were issued, plus over 80% of those rated 'amber'.[18] At the same DMC in 2019, 45% of 'African Visitor Visas' rated 'red' were refused compared to just 0.5% of those rated 'green'.[19]

Of course, an ADM system can make many more decisions within a given timeframe than a single human decision-maker. While this can be beneficial in that affected individuals may receive decisions more quickly, it also means that the negative impacts of a flawed ADM system are likely to be much greater.

f. **Lack of authorisation in law for use of these tools** – In order that public officials remain accountable to Parliament and to the public, any actions taken by public decision-makers must be authorised by law.[20] It is vital that: 'the manner in which executive functions will be carried out and to whom they are to be delegated is published, transparent and reliable'.[21]Where an automated system is used to make a decision there is a question mark over whether that decision-making is lawful if use of the system is not expressly provided for in law. For example, section 2 of the Social Security Act 1998 allows for any decision made by an officer to also be made 'by a computer for whose operation such an officer is responsible'. In the

[18] Independent Chief Inspector of Borders and Immigration, 'An inspection of entry clearance processing operations in Croydon and Istanbul: November 2016 – March 2017' (July 2017) at 3.7, 7.10 and 7.11, available at 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.

[19] Independent Chief Inspector of Borders and Immigration, 'An inspection of the Home Office's Network Consolidation Programme and the "onshoring" of visa processing and decision making to the UK September 2018 – August 2019' (February 2020) at 7.26, available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/863627/ICIBI_An_inspection_of_the_Home_Office_s_Network_Consolidation_Programme.pdf.

[20] A Le Sueur 'Robot Government: automated decision-making and its implications for parliament' in A Horne and A Le Sueur (eds) Parliament: Legislation and Accountability (Oxford: Hart Publishing, 2016).

[21] *R(Bridgerow Ltd) v Cheshire West and Chester Borough Council* [2014] EWHC 1187 (admin) at para 36.

case of *Khan Properties Ltd v The Commissioners for Her Majesty's Revenue & Customs*,[22] the First Tier Tribunal (Tax Chamber) found that use of an automated system to issue a penalty notice was unlawful as the law only provided for the decision to be made by an individual officer and not a computer. After the decision, Parliament passed section 103 of the Finance Act 2020 which states that anything capable of being done by an officer of HMRC can also be done by a computer. However other departments are using automated systems without a clear authorisation for such use in primary or secondary legislation. It is possible that these departments consider that no express authorisation is necessary because they are only using these systems as tools. If that is the case however, it is even more important that how they are using these tools, when such tools are being used in the decision-making process, and who retains ultimate decision-making control is publicly available information.

What measures could make the use of AI more transparent and explainable to the public?

13. The Cabinet Office is currently piloting an 'Algorithmic Transparency Standard' (ATS). The ATS asks public sector organisations across the UK to provide information about their algorithmic tools. It divides the information to be provided into Tier 1 and Tier 2. Tier 1 asks for high-level information about how and why the algorithm is being used. Tier 2 information is more technical and detailed. It asks for information about who owns and has responsibility for the algorithmic tool, including information about any external developers; what the tool is for and a description of its technical specifications (for example 'deep neural network'); how the tool affects decision making; lists and descriptions of the datasets used to train the model and the datasets the model is or will be deployed on; any impact assessments completed; and risks and mitigations.

[22] [2017] UKFTT 830 (TC)

14. Our view is that the ATS does not go far enough to meet a minimum viable standard of transparency. At present, it is not compulsory for public sector organisations to engage with the ATS. This does not appear likely to change in the near future. In its response to the consultation 'Data: a new direction', the Government stated that it "does not intend to take forward legislative change at this time", despite widespread support for compulsory transparency reporting, and, indeed, the Data Protection and Digital Information Bill did not include any such requirements.
15. Moreover, even if it were placed on a statutory footing, the ATS does not ask for sufficient operational detail. The purpose of transparency bears on its meaning in the context of ADM. The Berkman Klein Center conducted an analysis of 36 prominent AI policy documents to identify thematic trends in ethical standards.[23] They found there is convergence around a requirement for systems to be designed and implemented to allow for human oversight through the "translation of their operation into intelligible outputs". In other words, transparency requires the 'translation' of an operation undertaken by an ADM system into something that the average person can understand. Without this, there can be no democratic consensus-building or accountability. Another plank of meaningful transparency is the ability to test explanations of what an algorithmic tool is doing. In our view, meaningful transparency requires that people lacking specific technical expertise—i.e., the vast majority of us—are able to understand and test how an algorithmic tool works.[24]
16. Against this definition of transparency, the ATS falls short. Neither Tier 1 nor Tier 2 of the ATS requires sufficient operational details for individuals properly to understand the decision-making process to which they are subjected. At Tier 1, organisations are asked

[23] Jessica Fjeld, and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1 (15 January 2020), available at https://ssrn.com/abstract=3518482 or http://dx.doi.org/10.2139/ssrn.3518482.
[24] Mia Leslie and Tatiana Kazim, 'Executable versions: an argument for compulsory disclosure (part 1)' (The Digital Constitutionalist, 03 August 2022), available at: https://digi-con.org/executable-versions-part-one/.

to explain ‘how the tool works’, but nowhere is there a reference to any criteria or rules used by simpler algorithmic tools. At Tier 2, a ‘technical specification’ is requested, but this appears to mean nothing more than a brief descriptor of the type of system used, e.g., ‘deep neural network’.[25]

17. We propose the following measures for improving transparency:

    a. Requirements such as those within the ATS should be on a statutory footing. All public sector organisations should have statutory transparency obligations, in respect of all algorithmic tools involved in decision making with potential public effect. There will need to be careful consideration of how to make this new legislative scheme work alongside existing data protection laws and the Freedom of Information Act.

    b. ‘Public sector organisation’ should be broadly defined, along the lines of section 6(3)(b) of the HRA, under which a ‘public authority’ includes organisations whose functions are of a public nature.

    c. There should be no exemptions to the compulsory provision of information to the ATS. There should, however, be some limited exemptions to publication, the applicability of which falls to be determined not by the user of the algorithmic tool but by the operator of the ATS. These exemptions could be modelled on the FOIA and be subject to a consideration of the balance of the public interest in disclosure versus non-disclosure. Exemptions should not apply to the publication of high-level information, including the fact that there is an algorithmic tool in use in a particular context. If more detailed information about a particular tool is withheld on the basis of an exemption, there should be readily accessible avenues for challenging this,

[25] See also Tatiana Kazim and Sara Lomri, ‘Time to let in the light on the Government’s secret algorithms (Prospect, 02 March 2022), available at https://www.prospectmagazine.co.uk/politics/time-to-let-in-the-light-on-the-governments-secret-algorithms; and Mia Leslie and Tatiana Kazim, ‘Executable versions: an argument for compulsory disclosure (part 1)’ (The Digital Constitutionalist, 03 August 2022). Available at https://digi-con.org/executable-versions-part-one/.

with the possibility of review by an independent regulator such as the Information Commissioner's Office (ICO).

d. It should be compulsory for organisations to monitor the operation of the tool and update the information periodically and/or whenever there is a significant change.

e. The existing documents associated with the ATS should be redrafted to include clearer and more detailed guidance about the information to be provided.

f. At Tier 1, the high-level explanation of how the algorithm works should include rules or criteria used by the algorithm.

g. At Tier 2, there should be sufficient detail for an individual whose rights may be affected to fully understand how the process works. For example, we invite the Cabinet Office to consider whether a link to an 'executable version' (EV) should be a requirement. PLP has written in detail about disclosure of executable versions as a means to secure meaningful transparency.[26] On our definition, an EV of a model is one that allows someone with access to it to: (1) change the inputs or assumptions of the model; (2) run the model; and (3) see the outputs. In our view, the salience of an EV is that it allows someone to see and use the 'front-end' of the decision-making tool. It does not offer access to the 'back-end', and it is only a copy – it does not allow a third party to make changes to the system actually used by the decision-maker. Where an EV is disclosed, a third party or affected individual can run the EV on a range of different inputs and generate their own counterfactual explanations e.g., they would be able to say "If I earned more, my application would have been successful". Arguably, an EV is—for transparency purposes—the closest equivalent to a written policy because it allows an ordinary person to understand and test explanations of how discretionary power is to be exercised in a given case. For an example of an EV, the investigative newsroom Lighthouse

[26] Mia Leslie and Tatiana Kazim, 'Executable versions: an argument for compulsory disclosure (part 2)' (*The Digital Constitutionalist*, 03 November 2022), available at: https://digi-con.org/executable-versions-part-two/.

Reports, based in the Netherlands, were able to obtain sufficient information to create an EV by reconstructing the algorithm previously used by Dutch municipalities in an attempt to stop welfare fraud.[27] The algorithm profiles citizens on social assistance benefits and categorises them into risk groups, with specific target groups being rendered as potential fraudsters,[28] and Lighthouse Reports' EV allows people to generate their own risk score, understand how those scores are calculated, and what characteristics and behaviours are taken into account.

h. Alongside impact assessments, Tier 2 should include other ongoing monitoring and evaluation work conducted by Government departments, such as data on the impact and efficacy of their algorithmic tools.

To what extent is the legal framework for the use of AI, especially in making decisions, fit for purpose?

18. The current legal framework for the use of AI and ADM systems is fragmented. Few existing laws are AI-specific. Nonetheless, the existing legal framework contains a number of crucial safeguards which, even if they were not created with AI in mind, can be interpreted to regulate its use. Laws regulating the use of ADM systems include: the Freedom of Information Act 2000 (FOIA 2000); the Data Protection Act 2018 (DPA 2018) and UK General Data Protection Regulation (GDPR); public law doctrines developed through the common law, such as the duty to give reasons; the Equality Act 2010 (EA 2010); and the Human Rights Act 1998 (HRA 1998) and European Convention on Human Rights (ECHR), particularly articles 8 which protects the right to private and family life, home and correspondence and 14 which ensures equal enjoyment of all ECHR rights and

[27] FNV, Lighthouse Reports, Argos, and NRC news' reconstruction of the Dutch 'Fraud Scorecard' algorithm (14 July 2022), available at: https://www-fnv-nl.translate.goog/nieuwsbericht/sectornieuws/uitkeringsgerechtigden/2022/07/ben-jij-door-je-gemeente-mogelijk-als-fraudeur-aan?_x_tr_sl=auto&_x_tr_tl=en&_x_tr_hl=en-US&_x_tr_pto=wapp

[28] Gabriel Geiger and others, 'Junk Science Underpins Fraud Scores' (*Lighthouse Reports*, 22 June 2022), available at: https://www.lighthousereports.nl/investigation/junk-science-underpins-fraud-scores/.

protection from discrimination. In what follows, we give some additional detail on the requirements of the FOIA 2000, DPA 2018 and UK GDPR, and the EA 2010:

i. **FOIA 2000** – Section 31 of FOIA exempts the Government from requests where one of the exemptions from disclosure is engaged. Where an exemption is relied upon the likely prejudice to the state must outweigh the public interest in disclosure. Exemptions that are commonly relied upon in this context includes 31(1)(a) (disclosure is likely to prejudice the detection and prevention of crime) and s 31(1)(e) (disclosure likely to prejudice the operation of immigration controls). In principle s 31 strikes an acceptable balance of rights between the need of the state to effectively carry out its aims and the goals of open and transparent Government. However, it is PLP's experience, as noted earlier in this submission, that the s 31 exemption is being inappropriately relied on by Government departments. The DWP has relied on s 31(1)(a) when PLP has requested more information about an algorithm used to detect benefit fraud. The Home Office has relied on ss 31(1)(a) and (e) when PLP has tried to find out further information about the sham marriage algorithm. Given the literature described above indicates that the actual risk of 'gaming' or subverting the operation of the algorithm is minimal, the blanket refusal to disclose the factors used by these systems seems to be an inappropriate application of the exemption that requires further justification. Moreover, as we have also highlighted above, there must be further engagement with the public interest test and careful consideration of the public interest in disclosure of information explaining how an ADM system works.

j. **DPA 2018 and UK GDPR** – The DPA 2018 and UK GDPR are unusual in that some provisions are specifically tailored to the use of AI. They contain a number of crucial safeguards. While there is room for improvement, PLP's view is that these safeguards are essential and are, for the most part, fit for purpose. To the extent that reform is needed, it is to strengthen protection for the rights of data subjects, not to water it down. The key safeguards include:

i. **Data subject access requests** – are provided for under Article 15 of the UK GDPR. They allow individuals to check the accuracy of their personal data, learn more about how their data is being used and with whom their data is being shared, and obtain a copy of the data held about them. They are also an important investigative tool, helping to ensure transparency and uncover and prevent maleficence.

   We think it is vital that the data subject access requests are free of charge. This is because the ability of an individual to access their own data is a fundamental right. Moreover, some protected characteristics including race, sex (in the case of single mothers), and disability are associated with an increased risk of poverty.[29] The more protected characteristics someone holds, the greater their statistical risk of poverty.[30] As explained above, many known ADM systems, such as immigration enforcement systems and welfare fraud and error detection systems, appear to have a disproportionate effect on these groups. It is especially important that they have adequate, accessible options for finding out about the ADM systems to which they may have been subjected.

ii. **Article 22 of the UK GDPR** – provides that a 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." Article 22 – or something like it – is essential for ensuring human oversight in public decision-making. It also helps to ensure transparency: under Article 22, in combination with Articles 12, 13, and 14 (governing transparency of information) data subjects can

[29] See, for example, Sara Davies and David Collings, 'The Inequality of Poverty: Exploring the link between the poverty premium and protected characteristics' (February 2021), University of Bristol Personal Finance Research Centre, available at https://fairbydesign.com/wp-content/uploads/2021/02/The-Inequality-of-Poverty-Full-Report.pdf.
[30] Ibid.

get information to find out if they are being subjected to solely automated decision-making.

In our experience, it appears that Article 22 is capable of placing a meaningful limit on the deployment of ADM systems. For example, in response to PLP's request for information about Her Majesty's Prison and Probation Service's (HMPSS) use of algorithmic decision-making, a representative said: "I hope it is helpful if I explain that HMPPS does not, and will not use computer generated algorithms to automate or replace human decisions in HMPPS that have any significant impact on staff, people in prison or on probation. We may seek to automate low-level administrative decisions but this will always be deployed with human oversight and stringent quality assurance measures. We do not have any examples where 'a computer automatically performs all of the decision-making process without a human's direct involvement'".

This suggests that at least some Government departments may be interpreting Article 22 broadly, so as to meaningfully restrict the role of automation in decision-making.

Nonetheless, PLP considers that Article 22 could be made more effective through clarification of its key terms – especially "a decision based solely on automated processing" – to ensure that it has broad practical application. Article 22, properly defined, should prohibit *de facto* solely automated decision-making where, due to automation bias[31] or for any other reason, the human official is merely rubber-stamping a score, rating or categorisation determined by the computer. It should require meaningful human oversight, rather than a token gesture.

[31] See, for example, L.J. Skitka and others, 'Does automation bias decision-making?' (1999) 51 International Journal of Human-Computer Studies 991.

iii. **The requirement to undertake a Data Protection Impact Assessment (DPIA)** – is provided for under section 64 of the DPA 2018. Section 64(1) requires that a DPIA must be carried out "[w]here a type of processing is likely to result in a high risk to the rights and freedoms of individuals". Under section 64(3), the DPIA must include: "(a) a general description of the envisaged processing operations; (b) an assessment of the risks to the rights and freedoms of data subjects; (c) the measures envisaged to address those risks; (d) safeguards, security measures and mechanisms to ensure the protection of personal data… taking into account the rights and legitimate interests of the data subjects and other persons concerned." This requirement applies to all data controllers. PLP considers that DPIAs are an important tool for guarding against some of the risks posed by ADM systems, including discrimination. They can help with the identification and minimisation of such risks before they arise. This benefits providers and users of ADM technologies in that it helps them to avoid wasting resources on developing systems that flout legal standards protecting individual rights. It benefits individuals in that it helps to prevent them from being subject to rights-violating systems. In our experience, DPIAs appear to have been useful in helping Government departments to think through the full implications of using a given ADM technology and to help avoid legal challenges.

iv. **The requirement that data processing is lawful, fair and transparent** – is found is Article 5(1)(a) of the UK GDPR. The potential of this provision to constrain harmful and rights-violating uses of AI is evident in the recent action taken by the ICO against Clearview AI. In May 2022, the ICO fined Clearview AI Inc £7,552,800 for using images of people in the UK (and elsewhere) that were 'scraped' from the web and social media, without data subjects being made aware, and used to create a global online database

to support facial recognition. The ICO also issued an enforcement notice, ordering the company to stop obtaining and using the personal data of UK residents in this way, and to delete the data of UK residents from its systems. Amongst other provisions, the ICO relied on Article 5(1)(a) and, relatedly, Articles 6 and 9(2), which set out conditions for lawful processing of data and personal data, and Article 14, which requires data subjects to be informed about how their personal data is used.[32]

k. **EA 2010** – The Public Sector Equality Duty (PSED), set out in section 149 of the EA 2010 and elucidated by the Court of Appeal in *Bridges*,[33] is crucial in ensuring fairness for those subject to ADM. Section 149 requires public authorities to have due regard to equality considerations when exercising their functions. Meaningfully performed, the PSED helps public bodies ensure that decisions do not have unlawful discriminatory impact. Section 149 consolidated and replaced pre-existing specific duties concerning race, disability and sex. It extended coverage to the additional "protected characteristics" of age, gender reassignment, religion or belief, pregnancy and maternity, sexual orientation and, in certain circumstances, marriage and civil partnership. Under the EA 2010, it is the prerogative of the decision-maker to decide the process to undertake to ensure compliance with the PSED. In practice, public bodies often carry out Equality Impact Assessments (EIAs) in order to fulfil their obligations under section 149. Like DPIAs, EIAs are, in our experience, important in helping public bodies to avoid developing and deploying discriminatory AI tools and also a useful source of information, to help affected individuals understand the ADM systems to which they are subjected. As explained above, the *Bridges*[34] litigation concerned the South Wales Police's

[32] See 'ICO fines facial recognition database company Clearview AI Inc more than £7.5m and orders UK data to be deleted' (23 May 2022), https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/05/ico-fines-facial-recognition-database-company-clearview-ai-inc/.
[33] *R (Bridges) v Chief Constable of South Wales Police and others* [2020] EWCA Civ 1058.
[34] *R (Bridges) v Chief Constable of South Wales Police and others* [2020] EWCA Civ 1058.

use of facial recognition technology. Before the High Court, there was evidence that, due to imbalances in the representation of different groups in the training data, such technologies can be less accurate when it comes to recognising the faces of people of colour and women. The Court of Appeal found that the PSED imposes a duty on public bodies using ADM systems to take reasonable steps to gather relevant information about the impact of the system on people with protected characteristics. Public bodies must give advance consideration to issues of discrimination before making a decision to use a technology. They must make their own enquiries, rather than uncritically relying on the assurances of a third-party contractor. The PSED required the SWP to do "everything reasonable which could be done… in order to make sure that the software used does not have a racial or gender bias";[35] in other words, to ensure that the software was no worse at recognising the faces of people of colour or women.

However, the EA 2010 is not solely concerned with fairness on a statistical level but with the treatment of individuals. Dee Masters and Robin Allen KC have given the following example: "Suppose a situation in which a recruitment tool is used to identify 10 candidates for a particular role. There are 1000 applicants, 300 are men and 700 are women. "Outcome fairness" might be used to dictate that 30% of people identified as suitable candidates for the role must be men and 70% must be women meaning that the final recommended pool should consist of 3 men and 7 women… if there were 8 women who were most suitable, one woman would need to be "held back" so that 3 men could be put forward and the "right" statistical outcome achieved."[36] This example, judged by the standard of 'outcome fairness' alone, may seem unproblematic. But, under section 13 of the EA 2010, the woman

[35] *R (Bridges) v Chief Constable of South Wales Police and others* [2020] EWCA Civ 1058 at [201].
[36] Robin Allen KC and Dee Masters, Joint second opinion in the matter of the impact of the proposals within "Data: a new direction" on discrimination under the Equality Act 2010, available at https://research.thelegaleducationfoundation.org/wp-content/uploads/2021/11/TLEF-Second-Opinion-5-November-2021.pdf.

who is "held back" is subject to direct discrimination on the basis of sex. We should not lose sight of the unfairness of direct discrimination, simply because we are in an ADM context. In PLP's view, the courts and their interpretation of all the requirements of the EA 2010 have a valuable role to play in assessments of fairness in the context of ADM. We agree with Dee Masters and Robin Allen KC that the DPA 2018 and UK GDPR should not be "siloed" off from the EA 2010. We endorse their recommendation that the DPA 2018 and UK GDPR should be amended to state "unequivocally, and without any exceptions, that data processing which leads to breaches of the EA 2010 is unlawful".[37]

19. In summary, the existing legal framework provides a patchwork of vital, albeit imperfect, safeguards. If the legal framework were to be reformed, the focus should be on fortifying existing safeguards and ensuring clarity and coherence between existing laws.
20. Furthermore, it will be necessary to ensure that there are quick and effective ways of enforcing existing rights. At PLP's roundtables, one common view was that there is no need for new digital rights, but there is a need for effective enforcement mechanisms for affected individuals.

How should the use of AI be regulated, and which body or bodies should provide regulatory oversight?

21. As mentioned above, we think that a compulsory algorithmic transparency regime should be operated by an independent regulator, such as the ICO or, potentially, a dedicated AI regulatory body. If the ICO were to carry out this function, it would be important to ensure that it was capacitated with the necessary technical expertise and sufficient funding.
22. More generally, we think that AI regulation must be based on the following principles:

[37] Robin Allen QC and Dee Masters, Joint second opinion in the matter of the impact of the proposals within "Data: a new direction" on discrimination under the Equality Act 2010, available at https://research.thelegaleducationfoundation.org/wp-content/uploads/2021/11/TLEF-Second-Opinion-5-November-2021.pdf.

a. **Anti-discrimination** – Built-in bias and the risk of discrimination is, as explained above, a major concern when it comes to ADM. Existing practices for guarding against this risk include EIAs and DPIAs. These are important safeguards. However, as the *Bridges*[38] case showed, these assessments may not be carried out adequately or often enough. Our own investigative research tends to confirm this. We consider that there is a need for a more robust and proactive approach to guarding against the risk of discrimination. As well as considering the risk of discrimination at the outset, at the point of development and before roll-out, new technologies should be reviewed routinely throughout the period of deployment to check for indirect discrimination, and these evaluations should be made public to allows others to scrutinise the findings. Further, Government departments should not rely uncritically on assurances from third parties but must satisfy themselves that technologies they use are not discriminatory – even, or especially, where the system has been developed by a private contractor.

b. **Reflexivity** – In order to ensure that ADM systems are operating lawfully, fairly, and without bias, it is not sufficient to rely on external challenges. The Government department responsible for the system should be reflexive in its approach. In a research context, reflexivity refers to the examination of the researcher's own beliefs, judgments and practices and how these may have influenced the research. In the context of Government use of ADM systems, we envisage that reflexivity would involve proactively considering the ways in which the beliefs and judgments of people who developed the system may have influenced the way it works and taking action accordingly. For example, it will be important to consider the way that unconscious bias may affect the selection of training data and, therefore, the outputs of a machine learning algorithm. The dataset may need to be modified to mitigate this. Assuming there is a 'human in the loop', reflexivity would also involve

[38] *R (Bridges) v Chief Constable of South Wales Police and others* [2020] EWCA Civ 1058.

continuously reviewing the ways in which the beliefs and judgments of the officials may influence their approach to the ADM system's outputs. For example, if automation bias is identified as a risk, it may be necessary to provide training on and/or warnings about this. Reflexivity has overlaps with anti-discrimination: a reflexive approach would assist Government departments in effectively checking for risks of discrimination, both at the outset and throughout the period of deployment.

c. **Respect for privacy and data rights** – Respect for privacy and data rights must be central in the development and deployment of ADM technologies. Minimum safeguards would likely include adequate notice and opportunities for consent, as well as mechanisms allowing individuals to have continuing control over the use of their data. An adequate opportunity for consent means that there is a genuine choice available. For example, such consent cannot be a requirement for accessing essential services.

d. **Meaningful transparency** – Arguably, transparency has intrinsic value. But it also has instrumental value. It allows for proper debate and consensus-building around the use of new technologies in the public interest. And it is necessary in order for individuals and organisations to be able to hold the state to account and prevent maleficence. Various provisions of the UK GDPR impose duties that help to ensure a degree of transparency in relation to ADM systems. However, we consider that meaningful transparency requires more: not only a publicly available list of ADM systems in Government, but an adequate explanation of how they work. Executable versions of listed algorithms should also be available. Although there is ICO guidance on explaining decisions made using AI, it appears to us that this is not being followed consistently by Government departments. In our experience, it is often difficult to find out about the existence of an ADM system, let alone get an explanation of how it works – both in general and in application to a specific individual.

e. **Accountability and avenues for redress** – Accountability is to an extent dependent on transparency. But the two are not equivalent. Accountability goes beyond transparency, in that it requires adequate avenues for people to challenge the development and deployment of ADM systems, together with effective enforcement mechanisms and the possibility of sanctions. Definitions of accountability differ. But it has been suggested that any adequate definition will involve three elements: (i) Responsibility for actions and choices. There must be an accountable party who can be praised, blamed, and sanctioned; (ii) Answerability, which includes: first, capacity and willingness to reveal the reasons behind decisions to a selected counterpart (this could be the community as a whole); and, second, entitlement on the part of the counterpart to request that the reasons are revealed; (iii) Sanctionability of the accountable party, where 'sanctions' range from social opprobrium to legal remedies.

23. In practice, the realisation of these principles – especially the principles of anti-discrimination and respect for privacy and data rights – may require the prohibition of certain types or uses of AI technologies (see further below, 'What lessons, if any, can the UK learn from other countries on AI governance?' para.19).
24. Further, it will be vital to create quick and effective avenues for redress for affected individuals. One way to do this could be through a specialist regulator and forum for complaints relating to Government use of AI. However, roundtable attendees were concerned that a specialist forum may not be accessible for affected individuals. Another option could be sector- or system-specific avenues for redress. For example, if welfare benefits are suspended through the use of an automated system, an affected individual should be specifically informed that an automated system was used in the decision-making process and there should be a dedicated complaints procedure if they suspect unfairness.

What lessons, if any, can the UK learn from other countries on AI governance?

25. It is worth considering the EU Commission's proposed artificial intelligence (AI) regulation, adopted on 21 April 2021. The proposed regulation would include:
    - A blanket ban on certain AI systems that are considered to pose an 'unacceptable risk', including subliminal manipulative systems; systems which exploit vulnerabilities related to age, and physical or mental disability to distort behaviour; public sector 'social credit' systems; and real time remote biometric systems in public spaces.
    - A public register of 'high risk' AI systems in the form of a database, managed by the EU Commission, to which AI providers would be obliged to provide meaningful information about their systems;
    - A requirement that 'high risk' systems obtain certification indicating conformity to regulatory standards; and
    - A requirement that the design of 'high risk' AI systems allows for effective human oversight.
26. In our view, there is merit in some of these proposals. First, the proposal that the public register of AI systems is managed by the EU Commission, rather than by providers or users of the systems in question. As articulated above, we consider an independent regulator to be an effective aspect of a compulsory algorithmic transparency regime in the UK. Second, the proposal that the design of AI systems must allow for effective human oversight. If a similar principle was incorporated in the UK's regulation of AI, it would encourage a broader practical application of Article 22, that would prohibit *de facto* solely automated decision-making where, due to automation bias or for any other reason - requiring meaningful human oversight, rather than a token gesture. Third, the proposal to ban particularly problematic types or uses of AI to protect against specific harms to individuals and communities. However, we do not believe the UK should follow in devising an exhaustive list of prohibitions as an exhaustive list may allow for loopholes to develop or may quickly become outdated in light of new and unforeseen technological

developments. Many attendees to our roundtables considered that a more fruitful possibility may be the prohibition of harmful *uses*, rather than *types*, of new technologies.

27. Whilst offering many positive examples, is also important to acknowledge the limitations of the EU AI regulation proposal. For example, amongst attendees at our roundtable events, there was widespread concern about using 'high risk' as a touchstone for regulation and, in particular, as the touchstone for compulsory transparency. This is due to the difficulty of coming up with a satisfactory definition of 'high risk' and the potential for the threshold requirement to be manipulated or abused, such that AI systems that should be made transparent remain opaque. In the proposed EU regulations, 'high risk' is defined with reference to a list of functions, set out at Annex III. The list includes, AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons; AI systems intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending, or the risk for potential victims of criminal offences. This list can be added to by the EU Commission (by virtue of Article 7), but there is no "catch all" provision. The worry is that the list may quickly become outdated as new technologies and new uses of technologies emerge, and that it will be too slow and cumbersome to amend. Similar concerns were raised by attendees at our roundtables, with regard to a closed list of prohibitions. Such a list may allow for loopholes to develop or may quickly become outdated in light of new and unforeseen technological developments.
28. In Canada, the Directive on Automated Decision-Making (DADM) has been in effect since on 1 April 2019, requiring compliance as of 1 April 2020. The Directive seeks to make data and information on the use of ADM systems in federal institutions available to the public. The Directive creates a number of mandatory requirements:
    - Completing the Algorithmic Impact Assessment (AIA), a questionnaire requiring ADM operators to input details of the system. The AIA then produces a report that helps operators better understand and reduce the risks associated with the ADM system.

- Providing notice before decisions (through all service delivery channels to inform that the decision will be undertaken in whole or in part by an ADM system)
- Providing explanations before decisions
- The Government of Canada retains the right to access and test the ADM system and authorise external parties to review and audit these components as necessary – the person responsible for the ADM system must ensure that the software and related components 'are delivered to, and safeguarded by the department'
- Disclosure of source code
- Documenting decisions
- Testing data and information used by the ADM system for unintended data biases or other factors that may unfairly impact the outcomes (and developing processes to monitor the outcomes against these same standards on a scheduled basis)
- Ensuring human intervention
- Publish information on the effectiveness and efficiency of the ADM system in meeting programme objectives on a website or service

29. In our view, the DADM offers positive examples of AI governance in the public sector – particularly the compulsory requirement for all system operators to complete an algorithm-specific impact assessment and release the results on the Open Government Portal, in an accessible format and in both official languages. As set out above, we believe the UK ATS could be more effective if participation was compulsory, like that under the Canadian regime. The DADM also sets a positive standard in its requirement to inform individuals that a decision will be undertaken in whole or in part by an ADM system before the decision is made, and to provide explanations in advance of how a decision will be made. The requirement to test data and information used by the ADM system for unintended data biases or other factors that may unfairly impact the outcomes (and the requirement to continually monitor the outcomes against these same standards); and the requirement of human intervention in decision-making.

30. That being said, there are limitations to the Canadian regime. First, the Directive is very limited in scope – it only regulates systems in the Federal Government and federal agencies. It does not apply to systems used by provincial governments, municipalities, or provincial agencies such as police services, child welfare agencies and/or many other important public institutions. Nor does the Directive apply to private sector systems. Further still, requirements apply only to 'external services' of federal institutions. ADM systems used internally by federal institutions fall outside the scope of the DADM - a significant gap when one considers the expanding use of ADM systems in the employment context. The UK must consider this shortfall of the Canadian system to ensure that the scope of its own AI regulation is sufficiently broad and places a mandatory requirement on all public sector organisations.
31. Second, the Directive applies only to systems used to "recommend or make an administrative decision about a client". Canadian Professor, Teresa Scassa emphasised that "there may be many more choices/actions that do not formally qualify as decisions and that can have impacts on the lives of individuals or communities" which would fall outside the scope of the Directive and "remain without specific governance". The UK must take note of this limitation and ensure that its regulation of AI is not drafted so narrowly as to leave many forms of automation without specific governance.
32. Lessons may also by learnt by looking to France's Loi pour une Republique Numérique (Law for a Digital Republic) *2016.* [39] Under the mandatory regime of Loi pour une republique numérique, Article L-312-1-3 mandates transparency of all algorithms and ADM systems used by public agencies. France's Administrative Code is amended to include a right to an explanation of algorithmic decision-making. All agencies are required to publicly list any algorithmic tools they use, and to publish their rules, including systems where AI is only part of the final decision.

[39] Law No. 2016-321 of 7 October 2016 for a Digital Republic <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000033202746/> accessed 25 May 2022.

33. Like the Canadian DADM, the French regime requires administrations implementing ADM systems to provide notice that a decision is made or supported by an algorithm, but goes further by requiring this of all ADM systems in the scope of the Loi pour une Republique Numérique, not only those seen as high-risk. Under this regime it is also required to publish the rules the ADM system operates on, as well as the purpose of such processing.[40] Further still, if requested by the person concerned the implementing authority must also disclose the extent to which the algorithmic contributed to the decision-making process, the data processed and their sources, and the processing criteria and their weighting.[41]

Conclusion

34. To a large extent, the existing legal framework for the governance of AI is fit for purpose. It provides a patchwork of vital, albeit imperfect, safeguards across public law doctrines developed through the common law, such as the duty to give reasons. The law that makes up this patchwork includes: FOIA 2000, DPA 2018, UK GDPR, EA 2010, HRA 1998 and the ECHR, particularly Article 8, and Article 14. If the legal framework were to be reformed, the focus should be on fortifying existing safeguards, ensuring clarity and coherence between existing laws, and guaranteeing quick and effective ways of enforcing individual rights.
35. However, we consider the practical governance of AI, and specifically ADM, within this legal framework to be less effective. In our view, the law already requires transparency in Government use of automation. However, in practice, public bodies operating tools with a significant social impact, such as recommending who is to be investigated before they are allowed to marry, or whose benefits should be suspended, adopt an approach of secrecy by default.
36. While transparency is far from sufficient to secure the fair and lawful use of new technologies, PLP considers it to be a vital first step. Without transparency, there can be

[40] Article L.312-1-3, CRPA.
[41] Article R.311-3-1-2, CRPA.

no evaluation of whether systems are working reliably, efficiently, and lawfully, including assessment of whether or not they unlawfully discriminate. Without the necessary evaluations, there can be no accountability when automated decision making goes wrong or causes harm. Nor can there be democratic consensus-building about the legitimate use of new technologies.

37. In our view, securing meaningful transparency should be the first port of call when considering the regulation of AI. In this regard, inspiration could be drawn from other jurisdictions with compulsory transparency regimes, such as Canada and France.