Public Law Project (PLP) — Written evidence (NTL0046)

1. Public Law Project (PLP) is an independent national legal charity. For 30 years, PLP's mission has been to improve public decision-making and facilitate access to justice through a mixture of casework, advocacy and research. PLP's vision is a world in which individual rights are respected and public bodies act fairly and lawfully. Since October 2019, PLP's core focus area on public law and technology has specifically included automated decision-making (ADM) in government. Scrutinizing the use of algorithms and big data by government is now a key part of our work.
2. We are not opposed in principle to government use of ADM systems and we recognise the benefits of this. 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 that such systems operate fairly, lawfully and in a non-discriminatory way.

Introduction

3. We have found that government uses ADM in a range of areas, including but not limited to law enforcement. We know of two technologies of particular relevance to this call for evidence. We will refer to these as:
    a. The sham marriages algorithm; and
    b. The prisoner categorisation algorithm.

We discuss each of these in more detail below.

4. It is likely that there are many other instances of automation in the application of the law.

    A major challenge of working in this area is the lack of transparency around the existence, details and deployment of such systems. 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, which is unfortunate, especially given that other options for learning about government ADM systems are limited.
5. The lack of transparency makes it virtually impossible to provide a complete and universal account of the costs of using automation in law enforcement. That being said, we have found that there is a fairly consistent set of problems arising in relation to the algorithms we know about (one of which is this very opacity), and it is reasonable to suppose that these problems may be replicated in other contexts.
6. A note on terminology: by 'algorithm', we mean a set of rules for performing a task or solving a problem. This includes, but is not limited to, computer-operated algorithms. By 'automated decision-making', or 'ADM', we mean decision-making processes that are fully or partially undertaken by a computer-operated algorithm.

Known technologies

Sham marriages algorithm[1]

7. Documents obtained by PLP under the Freedom of Information Act 2000 (FOIA) indicate that the Home Office is using an automated triage system to determine whether a proposed marriage should be investigated as a 'sham': a marriage to avoid immigration law, rather than a marriage based on a genuine relationship.
8. The algorithm comes into play once a couple has given notice to be married. It is used if one party or both parties: a) is not a 'relevant national' – as of 01 July 2021, a 'relevant national' is someone who is not a British or Irish citizen, or a person with settled status or pre-settled status granted under the EU Settlement Scheme (EUSS), or a person with a decision pending on a pre-existing application for EUSS leave; b) lacks appropriate immigration status; or c) lacks a valid visa.
9. Using undisclosed criteria, the algorithm sorts couples into 'red' and 'green' categories. A red light indicates that an investigation is required to identify or rule out sham activity.
10. A graph included in the Home Office's Equality Impact Assessment (EIA) shows that couples of Bulgarian, Greek, Romanian, and Albanian nationality are given a red light at a rate of between 20% and 25% – higher than the rate for any other nationality.
11. Couples flagged for investigation are asked to provide more information and, often, to attend an interview and cooperate with home visits. If they refuse, they will not be allowed to marry. If they do comply, immigration officers will use the new information to determine if the marriage is a sham. If the decision goes against the couple, they can still marry, but their immigration status will be at risk and one or both parties may face forced removal.

Prisoner categorisation algorithm[2]

12. In May 2019, the Ministry of Justice (MoJ) began piloting a system for the security categorisation of prisoners, known as the Digital Categorisation Service (DCS). The pilot covered nine prisons: HMYOI Aylesbury, HMP Belmarsh, HMP Elmley, HMP Lowdham Grange, HMP Maidstone, HMP Pentonville, HMP/YOI Rochester, HMP Standford Hill and HMP Thameside. In January 2020, the MoJ made plans to roll out the DCS in Category B, C and D prisons across the male prison estate.
13. One feature of the DCS is a computer-operated algorithm for calculating a prisoner's 'provisional category'. Documents obtained by PLP under the FOIA indicate that this algorithm works as follows: the DCS contains a list

[1] We have written about the sham marriages algorithm in an article for Free Movement, 'Home Office refuses to explain secret sham marriage algorithm' (21 July 2021), available at https://www.freemovement.org.uk/home-office-refuses-to-disclose-inner-workings-of-sham-marriage- algorithm/.

[2] We have written about the prison categorisation algorithm in an article for Inside Time, 'Security categorisation issues' (15 February 2021), available at https://insidetime.org/security-categorisation- issues/.

of newly sentenced offenders in need of initial categorisation, drawn from the Prison National Offender Management Information System (P-NOMIS). The prison officer selects a prisoner, and the DCS automatically populates an online form with information about the prisoner, also drawn from the P-NOMIS. The officer can add further information as required. The algorithm uses all this information to generate a 'provisional category' for the prisoner, which the officer can accept or reject.[3]

14. The DCS then generates a completed categorisation form. This form is supposed to provide information about the reasons for the decision, 'while not disclosing sensitive details of sources' and should be available to the relevant prisoner on request.

Costs of new technologies

15. In general terms, ADM technologies may have various benefits: saving time; reducing expenditure; and improving the quality and consistency of decisions. For example, the use of ADM to decide most applications under the EU Settlement Scheme through a check of existing Department of Work and Pensions and HM Revenue and Customs data systems enabled more than six million applications to be processed and, in many cases, positive decisions to be issued swiftly, reducing uncertainty and expense for both applicants and the state.

16. However, ADM technologies also come with significant costs. These include:

a. **Lack of transparency and accountability** – A common problem is opacity, whether intentional or due to the complexity of the system. The latter is a particular problem when it comes to machine learning. A machine learning algorithm may be a 'black box', even to an expert.

   Intentional opacity can occur where government, or a private developer contracted by government, deliberately withholds information about the system due to concerns about commercial confidentiality or possible abuse and circumvention of the system's rules.

   Opacity is arguably a cost in and of itself, but also comes with costs in terms of the ability of people to hold the state to account. We consider that transparency is a prerequisite for accountability (see further below).

b. **Privacy and data protection costs** – These new technologies generally involve the processing of data on a large scale. This comes with widely recognised costs when it comes to privacy and data protection[4]. In October 2019, the UN Special 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

---

[3] For more detail on the different prison categories and the categorisation criteria, see the Ministry of Justice and HM Prison and Probation Service, 'Security Categorisation Policy Framework (re-issued 17 August 2021), available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/10 11502/security-categorisation-pf.pdf.

[4] 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.

privacy and data protection. 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.[5]

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. [6]

c. **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.

The possibility of problems with the training data was highlighted in the well-known *Bridges* 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 BAME people and women[7]. 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.[8]

There can also be bias in the design of an algorithm's rules. Following a legal challenge mounted by JCWI and Foxglove, the Home Office was recently forced to scrap its visa streaming algorithm, 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. This, it was argued, amounted to racial discrimination and was a breach of the Equality Act 2010.

d. **Risk of unlawfulness and unfairness** – A number of common problems with ADM systems give rise to a risk of unlawful and/or unfair decision making. These problems include:

i. Mismatch between a system's overall purpose and its outputs. In

[5] *NJCM v The Netherlands* C-09-550982-HA ZA 18-388.
[6] 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_A dvanceEditedVersion.docx.
[7] 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.
[8] *R(Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058, [201].

social domains, an algorithm may make use of proxies. For example, it may use past events to predict the future. This could lead to inaccuracy and/or unlawfulness.

ii. Errors in the system's outputs (false positives, false negatives, or a combination of both).

iii. Feedback loops, meaning that the system's decisions simply reflect the decisions it has made in the past, rather than the actual state of the world. This could render decision unfair and/or unlawful.

iv. Inflexibility, arising from the fact that ADM systems generally work by applying fixed rules uniformly across a broad range of cases. Unlike a human decision-maker, an algorithm cannot make an exception for a novel case – it can only act in accordance with its programming.

v. Automation bias: a well-established psychological phenomenon whereby people put too much trust in computers[9]. This may mean that officials over-rely on automated decision support systems and fail to exercise meaningful review of an algorithm's outputs.[10]

17. Of course, an ADM system can make many more decisions within a given timeframe than a single human decision-maker. While this is a significant benefit of using an ADM system, the costs of flawed ADM systems are likely to be much greater.

18. As we currently understand it, several of these potential problems and associated costs may arise in relation to the sham marriages algorithm. We are aware of the following issues:

- Opacity around the criteria applied by the algorithm, potentially undermining the procedural fairness of decisions to investigate. In response to our original request for information, the Home Office refused to disclose the criteria used by the algorithm, and PLP requested an internal review. The Home Office responded in June 2021, insisting that publication of the criteria used to determine which couples should be investigated would be likely to prejudice the Home Office's ability to do so, and would not be in the public interest. PLP has asked the Information Commissioner's Office (ICO) to look at whether this refusal was justified.
- Potential discrimination. Some nationalities – including Bulgarian, Greek, Romanian, and Albanian people – are statistically more likely to be targeted for investigation than others. Even if the algorithm's criteria do not include nationality, it may nonetheless be indirectly discriminatory if it is having a systemic negative

[9] See, for example, L.J. Skitka and others, 'Does automation bias decision-making?'(1999) 51 International Journal of Human-Computer Studies 991. For an example of automation bias in action in the UK, see the 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 An-inspection-of-entry-clearance-processing-operations-in- Croydon-and-Istanbul1.pdf (publishing.service.gov.uk).

[10] See also our response to the Law Commission's consultation on their 14th Programme of Law Reform, in which we outline some of the legal issues that arise as a result of these problems, available at Law-Commission-Consultation-Response-FINAL.pdf (publiclawproject.org.uk).

impact on people of these nationalities, contrary to sections 19 and 29 of the Equality Act 2010. The Home Office provided a redacted Equality Impact Assessment (EIA), including an incompletely labelled graph. The EIA stated that '[a] review of the nationalities involved has been conducted', but this fuller analysis was not provided. It is not clear that the Home Office has done 'all they reasonably could' to guard against the risk of discrimination.

- Possible automation bias. If the official conducting a sham marriage investigation is aware that the couple has been given a red light by the algorithm, they may be unduly inclined to conclude that the relationship is a sham. This may amount to a fettering of the official's discretion or reliance on irrelevant considerations and could mean that couples unfairly face an adverse finding following investigation, with potentially severe consequences for their relationship and immigration status. It is unclear what if any measures have been put in place by the Home Office to mitigate this risk, such as training officers to be alert to automation bias or including a warning when officials are notified of a red light.

19. Similar problems may arise in relation to the prisoner categorisation algorithm, including:

    - Opacity. Here again, little is known about the rules applied by the algorithm. The MoJ has provided no information about the rules, apart from stating in a Data Protection Impact Assessment (DPIA) that it uses 'approved data decision trees'. Further, it appears from the DPIA that prisoners may not be given enough information to understand the decision to place them in a particular category, and to challenge the decision if they think it is wrong.
    - Potential discrimination. The algorithm may be over-categorising BAME prisoners.

      During the DCS pilot, BAME prisoners were initially categorised as Category B at a higher rate than white British prisoners: 10% as compared to 7%.

    - Possible automation bias. The MoJ's pilot evaluation noted that officers tended to over- rely on the provisional category, without assessing whether it was the most appropriate one. This mainly occurred in cases where the algorithm recommended Category C, when Category D could have been considered. During the pilot, 95.7% of initial categorisations were accepted by the categoriser. In most of the remaining cases, the categoriser intervened to increase the prisoner's category. This could undermine the fairness of ultimate categorisation decision.

20. Finally, we note that, in the context of law enforcement, the implications of an automated decision are likely to be particularly acute:

    - Media reports indicate that a sham marriage investigation can be intrusive and harmful, regardless of its outcome.[11] If the couple

[11] See, for example, Diane Taylor and Francis Perraudin's article in the Guardian, 'Couples

does not comply with the investigation, they will not be allowed to marry. And if the decision following investigation goes against them, this can have serious consequences for their right to remain in the country.

- Very significant consequences flow from prison categorisation, too. For example, higher risk prisoners generally have less freedom in prison than other prisoners and less access to programs and activities.

21. In short, an unfair or unlawful adverse decision in both of these contexts has the potential to inflict significant harm on those it affects.

Possibilities for a new legal framework

22. It is worth considering the EU Commission's proposed artificial intelligence (AI) regulation, adopted on 21 April 2021[12]. The proposed regulation would include:

    - **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;
    - **Certification** indicating conformity to regulatory standards;
    - A requirement that the design of high risk AI systems allows for **effective human oversight**; and
    - **A blanket ban on certain AI systems**, 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.

23. PLP considers that the UK should be exploring similar options. In particular, we think there may be value in a public register of high risk ADM systems in the UK – though the efficacy of such a measure would depend on a number of factors. In particular, it would depend on: what information is provided (see below at paragraph 24(d), 'Meaningful transparency'); the extent of the duty to provide this information; and the threshold of 'high risk' which, in our provisional view, should be interpreted broadly.

Guiding principles

24. We recommend that guiding principles should include:

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

---

face 'insulting' checks in sham marriage crackdown' (14 April 2019), available at https://www.theguardian.com/uk-news/2019/apr/14/couples-sham-marriage-crackdown-hostile- environment.

[12] The proposal in available at EUR-Lex - 52021PC0206 - EN - EUR-Lex (europa.eu).

against this risk include EIAs and DPIAs. These are important safeguards. However, as the *Bridges*[13] 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, new technologies should be reviewed routinely throughout the period of deployment to check for indirect discrimination.

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 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

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

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.

Articles 12 ('Transparent information, communication and modalities for the exercise of the rights of the data subject') and 22 ('Automated individual decision- making, including profiling') of the General Data Protection Regulation 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.[14] Executable versions of listed algorithms should also be available.

Although there is ICO guidance on explaining decisions made using AI,[15] 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 getting an explanation of how it works – both in general and in application to a specific individual.

e. Accountability

Accountability is to an extent dependent on transparency.[16] 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

[14] For an account of what it means to adequately explain a decision made using AI, see ICO's guidance, 'Explaining decisions made with AI', available at https://ico.org.uk/for-organisations/guide- to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/.

[15] Available at https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining- decisions-made-with-artificial-intelligence/.

[16] Michele Loi, at the University of Zurich, and Matthias Spielkamp, at Algorithm Watch, helpfully analyse the relationship between transparency and accountability in their paper 'Towards accountability in the use of Artificial Intelligence for Public Administrations' (21 July 2021), available at https://algorithmwatch.org/en/wp-content/uploads/2021/05/Accountability-in-the-use-of-AI-for-Public- Administrations-AlgorithmWatch-2021.pdf.

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.[17]

We consider that this a useful starting point in coming up with a robust principle of accountability in the context of ADM and other new technologies.

Conclusion

25. While there may be benefits of using ADM in law enforcement, it also comes with significant costs. These costs could be reduced if the UK adopts principles such as those we have recommended, and begins to explore options along the lines of the EU Commission's proposed AI regulation. Otherwise, there is a real concern that the use of ADM in law enforcement could be unfair and discriminatory, and not in the public interest.

29 September 2021

[17] Ibid, at page 8.