Algorithms in the Criminal Justice System

A report by

The Law Society Commission on the Use of Algorithms in the Justice System

The Law Society of England and Wales

June 2019

Commission creator and director: Sophia Adams Bhatti

Commissioners:
Prof Sylvie Delacroix
Prof Sofia Olhede
Christina Blacklaws

Guest Commissioners:
Andrea Coomber, JUSTICE
Sir William Blair, Queen Mary University of London
Madhumita Murgia, *Financial Times*

Lead Author: Dr Michael Veale

Project Team:
Olivier Roth – Project Lead
Rebecca Veitch
William McSweeney
Catherine O'Gallachoir
Bal Matharu

2019, The Law Society of England and Wales

1 Table of Contents

1 Table of Contents 2

2 Executive Summary and Recommendations 4

2.1 List of Recommendations of the Commission 5

3 Scope of the Commission 9

4 Defining Algorithms 10

5 Introduction 12

5.1 Algorithms in Criminal Justice 12
5.2 Drivers 13

5.2.1 Resourcing pressures 13
5.2.2 Desires for increased proactivity 13
5.2.3 Data and infrastructure availability 14
5.2.4 New crime challenges 14
5.2.5 Access to justice 14
5.2.6 Belief in the efficacy of computational prediction 15

5.3 Benefits 15

5.3.1 Efficiency from automation 15
5.3.2 Efficacy from augmentation 16
5.3.3 Auditability 16
5.3.4 Consistency and control 16
5.3.5 Monitoring performance 17

5.4 Dangers 17

5.4.1 Instrumental concerns 17
5.4.2 Dignitary concerns 19
5.4.3 Justificatory concerns 21
5.4.4 Systemic concerns 22

6 Data protection and algorithms in criminal justice 25

6.1 Automated decision making for law enforcement purposes 27

6.1.1 Risk of rubber-stamping 27
6.1.2 Omission of discrimination provisions in DPA 2018 Part 3 28
6.1.3 Limited protection of groups 30
6.1.4 Omissions of transparency provisions in DPA 2018 Part 3 30

6.2 Unclear provisions around public-private partnerships 31
6.3 Data Protection Impact Assessments 32
6.4 Further issues 32

7 Algorithms in Criminal Justice Today 33

7.1 Predictive Hotspot Policing 33

7.1.1 Concerns 35

7.2 Facial Recognition in Policing 36

7.2.1 Concerns and legislation around data and privacy 39

7.2.2 Concerns and legislation around bias....42
7.2.3 Concerns and legislation around accountability, transparency and oversight....43
7.3 Individual Risk Assessment....45
7.3.1 Scoring at arrest....45
7.3.2 Scoring suspects....47
7.3.3 Scoring victims....47
7.3.4 Scoring staff....48
7.3.5 Scoring in the prison system....48
7.3.6 Concerns and legislative framework....53
7.4 Digital Forensics....55
7.5 Modernisation of the Courts....57
7.6 Mobile Device Extraction....58
8 Cross-cutting Rights and Issues....61
8.1 Lawfulness....61
8.2 Compliance Capacity....62
8.3 Transparency, Explanation and Justification....64
8.4 Discrimination and Computational Power....67
8.5 Concluding remarks....70
9 Annexes....72
9.1 Engaged Stakeholders....72
9.2 Lines of Enquiry....73
References....74

2 Executive Summary and Recommendations

The Law Society established the Technology and the Law Policy Commission to examine the use of algorithms in the justice system of England and Wales. The Commission considered both simpler ‘hand-crafted’ systems and more complex, computationally generated ones such as machine learning. It held four public evidentiary sessions, interviewed over 75 experts, and read over 82 submissions of evidence and many more supplementary studies, reports and documents on the topic.

This report contains findings and recommendations concerning the use of algorithmic systems in the **criminal justice** system. The Commission considered a range of currently deployed systems that fell within this brief, including individual risk assessment and recidivism prediction; prospective crime mapping and hot-spotting; and mobile phone data extraction tools.

At the most basic level, the Commission has found **a lack of explicit standards, best practice, and openness or transparency about the use of algorithmic systems in criminal justice across England and Wales**. This was concerning, as the high-stakes decisions and measures taken in the justice system demand extremely careful deployment. There are significant challenges of bias and discrimination, opacity and due process, consistency, amenability to scrutiny, effectiveness, disregard of qualitative and contextual factors, against a backdrop of the potential of these systems to more deeply change the nature of the evolution of the law. The Commission recommends that a **National Register of Algorithmic Systems should be created as a crucial initial scaffold for further openness, cross-sector learning and scrutiny**.

While many deployments are in a pilot or experimental stage, the Commission notes that the technologies being deployed are not so technically novel that they cannot be critically assessed by multi-disciplinary teams for their effectiveness, their conformity to real challenges, and their potential for unintended and undesirable side effects, particularly from optimising for some goals or aspects of an issue to the detriment of others. It is key **that in-house capacity is built and retained for overseeing and steering these systems, and that coordination occurs across the justice system to ensure this capacity is world-class**.

In-house capacity is only one piece of the puzzle. Governing algorithmic systems in criminal justice usually brings multi-dimensional tensions and value-laden choices to grapple. These **tensions emerge at many different points in development, deployment and maintenance, and are usually not between a ‘bad’ and a ‘good’ outcome, but between different values that are societally held to be of similar importance**. It is insufficient and unacceptable for the bodies and agencies involved to make these decisions alone, requiring instead the **engagement of broad stakeholders including civil society, academia, technology firms and the justice system more broadly**. Risk of systems being gamed is real, but often overstated in relation to the risks from lack of openness, engagement, and the loss of trust in procedural justice and the rule of law. Such risks stem especially from what are effectively policy decisions baked into algorithmic systems being made invisibly and unaccountably by contractors and vendors. The Commission’s work has highlighted that such crucial, often **political design choices should never be outsourced**.

The Commission has also analysed the broader and often new legal framework that in part governs algorithmic systems in criminal justice. In the course of evidence-taking, the Commission became heavily concerned that **some systems and databases operating**

today, such as facial recognition in policing or some uses of mobile device extraction, lack a clear and explicit lawful basis. This must be urgently examined, publicly clarified and rectified if necessary. While the United Kingdom has more explicit provisions covering algorithmic systems than many other parts of the world, these contain significant omissions and loopholes that need joined-up consideration. **The Commission recommends several clarifications and changes to data protection legislation, procurement codes, freedom of information law, equality duties and statutory oversight and scrutiny bodies** which would provide key safeguards to the integrity of criminal justice in the digital age.

Many of the **heavily individualised, legal safeguards proposed to algorithmic systems in commercial domains, such as individual explanation rights, are unlikely to be very helpful in criminal justice, where imbalances of power can be extreme and are exacerbated by dwindling levels of legal aid**. Societal, systemic oversight must be placed at the forefront of algorithmic systems in this sector, which will require innovative and world-leading policies. **The United Kingdom has a window of opportunity to become a beacon for a justice system trusted to use technology well, with a social licence to operate and in line with the values and human rights underpinning criminal justice. It must take proactive steps to seize that window now**.

2.1 List of Recommendations of the Commission

The Commission's recommendations are intended to provide a multi-faceted response to the issues identified during its investigations. There is no silver bullet, no single overarching policy or responsible body, but rather a deep need to create a cradle of responses which span vertically up and down the supply chain, and horizontally across the various agencies and actors. Some of these recommendations will take little to enable beyond a commitment, some might require policy or statutory change, while others will require cross-disciplinary and multi-stakeholder elaboration to create the right detail.

Thematically the recommendations fall into six groups: the need for improved oversight; the importance of strengthening algorithmic protections in data protection; going beyond data protection to enhance equality and human rights duties; baking in values and protection in procurement, design and purchasing; clarifying and respecting the law around the use of algorithmic systems; and plugging the analytical capacity gap.

Recommendation 1. Oversight – A range of new mechanisms and institutional arrangements should be created and enhanced to improve oversight of algorithms in the criminal justice system.

Sub-Recommendation 1.1 Sunset Clauses – Any future statutory requirements which require or encourage the use of algorithmic systems in criminal justice should be subject to sunset clauses requiring their automatic, full qualitative review.

Sub-Recommendation 1.2 Capacity of the Information Commissioner – The Information Commissioner must be adequately resourced to examine algorithmic systems with rigour on a proactive, rather than predominantly reactive basis.

Sub-Recommendation 1.3 Code of Practice for Algorithmic Systems in Criminal Justice – The Government should request and resource the Information Commissioner to create a code of practice for algorithmic systems in criminal justice under the Data Protection Act 2018 s128(1).

Sub-Recommendation 1.4 Centre for Data Ethics and Innovation – The Centre for Data Ethics and Innovation should be given statutory footing as an independent, parliamentary body, with a statutory responsibility for examining and reporting on the capacity for public bodies, including those in criminal justice, to analyse and address emerging challenges around data and society in their work, and develop a taxonomy of concepts important to algorithmic systems across sectors and domains.

Sub-Recommendation 1.5 Super-complaints – The Government should make provisions for Article 80(2) of the GDPR, which allows civil society organisations to complain to the ICO and seek a judicial remedy on behalf of a group rather than an individual. This provision should apply to the whole Data Protection Act 2018, including Part 3, rather than just the GDPR.

Sub-Recommendation 1.6 Public Interest Access – A facility should be established to enable secure access to algorithmic systems in use by or on behalf of public bodies in the criminal justice system for researcher and journalistic oversight. The British Library and the Centre for Data Ethics and Innovation could be candidates for coordinating this effort.

Sub-Recommendation 1.7 National Register of Algorithmic Systems – A register of algorithmic systems in criminal justice should be created, including those not using personal data, alongside standardised metadata concerning both their characteristics, such as transparency and discrimination audits and relevant standard operating procedures, and the datasets used to train and test them. Leadership of this could be taken by the Centre for Data Ethics and Innovation, as the Centre matures, in an open consultation procedure considering the criteria and thresholds for systems included in this register.

Recommendation 2. Strengthening Algorithmic Protections in Data Protection – The protections concerning algorithmic systems in Part 3 of the Data Protection Act 2018 should be clarified and strengthened.

Sub-Recommendation 2.1 Transparency Rights – The transparency provisions concerning profiling and algorithmic decision-making in the GDPR (particularly Articles 13(2)(f), 14(2)(g), 15(1)(h)) should be mirrored for law enforcement in Part 3 of the Data Protection Act 2018 (s 44–45) and subject to the same balancing test for disclosure, rather than removed entirely.

Sub-Recommendation 2.2 ICO Guidance on Logging for Algorithmic Systems – The ICO should provide guidance on how the logging requirements in Part 3 of the Data Protection Act apply to the use of algorithmic systems falling under this Part.

Sub-Recommendation 2.3 Data Protection Impact Assessments – Where Freedom of Information tests restrict release, a bespoke public-facing version of a data protection impact assessment concerning a consequential algorithmic system in criminal justice should be proactively published.

Sub-Recommendation 2.4 Meaningful Human Intervention – The Data Protection Act 2018 should be amended to specify the nature of the input needed to not to be a decision "based solely on automated processing" and trigger Article 22 of the GDPR, Section 14 and 49 of the Data Protection Act 2018.

Sub-Recommendation 2.5 Public Private Partnerships – The ICO should provide guidance on how the Data Protection Act Part 3 functions in the contexts of public-private partnerships and algorithmic systems.

Sub-Recommendation 2.6 Discrimination Provisions in Data Protection – The Government should explicitly transpose Article 11(3) of the Law Enforcement Directive concerning the prohibition on discrimination of algorithmic systems, and make explicit statutory provisions for ensuring that Part 2 ADM are not discriminatory under the powers to derogate from the GDPR provided by Article 22(2)(b), GDPR.

Sub-Recommendation 2.7 Significant Decisions and Groups – The Data Protection Act 2018 should be amended to clarify that a decision can be considered 'significant' if it impacts upon a protected or otherwise salient group to which a natural person belongs, rather than considering only impacts upon a single individual.

Recommendation 3. Protection beyond Data Protection – Existing regulations concerning fairness and transparency of activities in the justice sector should be strengthened in relation to algorithmic systems.

Sub-Recommendation 3.1 Public Sector Equality Duty – Given the importance of countering discrimination within algorithmic systems, Equality Impact Assessments should be formalised as a requirement before deploying any consequential algorithmic system in the public sector and these should be made proactively, publicly available.

Sub-Recommendation 3.2 Socioeconomic Equality Duty – Given algorithmic systems' high potential for socioeconomic discrimination, the Government should commence the socioeconomic equality duty in the Equality Act 2010 s1 in England and Wales, at least with regard to algorithmic decision-support systems.

Sub-Recommendation 3.3 Information Rights around Algorithmic Systems – The Government and/or Information Commissioner should provide guidance on how Freedom of Information Rights apply to value-laden, algorithmic software systems, particularly in the criminal justice sector.

Recommendation 4. Procurement – Algorithmic systems in the criminal justice system must allow for maximal control, amendment and public-facing transparency, and be tested and monitored for relevant human rights considerations.

Sub-Recommendation 4.1 Value-laden Decisions and Outsourcing – Value-laden decisions, such as problem definition, structuring, or choice between trade-offs in models, should never be explicitly or implicitly outsourced, for example through contracting or procurement.

Sub-Recommendation 4.2 Human Rights by Design – The Government should commission a review into policy options for mandating human rights considerations in technological design within different consequential sectors, including in the criminal justice system. This review should consider how and where human rights impact assessments should be required in public procurement processes.

Sub-Recommendation 4.3 Statutory Procurement Code – A procurement code for algorithmic systems in criminal justice should be developed, and a duty for relevant actors to adhere to it made a binding statutory requirement with a credible enforcement mechanism.

Sub-Recommendation 4.4 Individual Explanation Facilities and Remedies – Algorithmic systems in criminal justice must have explanation facilities focused on each decision or measure, designed to help individuals and users assess whether a given output is justified, and whether they should seek a remedy through the courts.

Sub-Recommendation 4.5 Societal Explanation Facilities – Algorithmic systems in criminal justice must have explanation facilities designed to allow broader internal and external scrutiny, such as over the general logics, functioning, behaviour and impact of the models.

Recommendation 5. Lawfulness – The lawful basis of all algorithmic systems in the criminal justice system must be clear and explicitly declared in advance.

Sub-Recommendation 5.1 Facial Recognition Model Use – Facial recognition systems must operate clearly under the rule of law, with their lawful basis explicitly and openly defined, and this assessment should be made publicly available.

Sub-Recommendation 5.2 Facial Recognition Datasets – Datasets used in facial recognition must operate clearly under the rule of law, adhering to conditions of strict necessity, and with categories of individuals clearly split as required under Part 3 of the Data Protection Act 2018. These must also specify how the data set has been selected to avoid selective sampling of the population, which could lead to bias and discrimination.

Sub-Recommendation 5.3 Biometrics Commissioner – The scrutiny powers, resources, and consultation role of the Biometrics Commissioner should be strengthened, and the scope of the Commissioner broadened and regularly reviewed.

Sub-Recommendation 5.4 Mobile Device Extraction Assessment – An appropriate body – potentially Her Majesty’s Inspectorate of Constabulary and Fire and Rescue Services (HMICFRS) – should be tasked with establishing a working group to consider issues around the legal, effective and legitimate use of technologies to search seized electronic devices.

Recommendation 6. Analytical Capacity and Capability – Significant investment must be carried out to support the ability of public bodies to understand the appropriateness of algorithmic systems, and where appropriate, how to deploy them responsibly.

Sub-Recommendation 6.1 Formalise Governance of Risk Scoring – The Government should take stock of the practices surrounding the development of risk assessment tools used in sentencing and offender management, and enshrine at least the current best practices – such as regular analysis, reviewing and reporting – as statutory responsibilities.

Sub-Recommendation 6.2 Research Support – The Government should support joint research projects between universities and actors in the justice sector around applied algorithmic systems, including how algorithmic analysis can promote equity in and access to justice.

Sub-Recommendation 6.3 Training Support – The Government should support universities in offering educational programmes for public interest practitioners sitting at the intersection of technology, law and human rights.

Sub-Recommendation 6.4 Digital Forensics In-House Capacity – The Government must ensure that the public sector maintains significant, effective capacity to rigorously understand digital forensic issues.

3 Scope of the Commission

The Law Society of England and Wales has a strong interest in the impact of technology and data use on human rights. Algorithms play an increasing role in all aspects of society and have caused and highlighted a range of concerns including unfairness, discrimination and opacity. But equally, technology – deployed with the right approach and careful consideration of the problems it is applied to, and underpinned by the principles of the rule of law – has the potential to have a net positive benefit to individuals, communities and society. Recent events have indicated that this outcome should not be taken for granted. Market forces or technology enhancements alone are unlikely to deliver the vision of the future hoped for without care and societal steer. The speed of scientific discovery, technological innovation and data capabilities pose new ethical, legal and social issues, and it is incumbent upon those in the relevant fields to explore and understand them and to develop technologies with human rights and flourishing at the core, rather than pushing on regardless.

The justice system is a particularly acute environment to analyse the impacts of technology. Wrong decisions threaten human rights and could go unrecognised, resulting in public trust in society being undermined. The justice system as a whole, and the values we have come to associate with it, might be transformed beyond recognition by uncritical reliance on data-intensive technologies, and that transformation comes with high individual and collective stakes.[1]

The Law Society called for contributions from all interested parties on the topic of algorithms in the justice system, in the form of oral evidence at public evidence sessions, one to one in depth interviews and written submissions. We were looking to hear from practitioners, academics, tech professionals, civil liberties organisations, companies that make algorithmic systems and tools, public bodies that use them, and anyone who has an interest in technology, the rule of law and human rights. It was our mission to build a collaborative and multi-disciplinary programme, to create in partnership with all those in the field a better understanding of the issues at play.

To help us understand this very broad issue, we set the following objectives of the Commission:

1. Investigate the current and future uses of rule-based systems and advanced algorithmic systems, including machine learning, in the English and Welsh justice system.
2. Identify and analyse ethical and human rights principles against computational possibilities.
3. Provide recommendations to those developing or using technologies in the English and Welsh justice system to successfully identify, prevent and mitigate harms.

Given the scope of this challenge, this report focuses on the criminal justice system. Decisions in this area have such significant impact on individuals that the distribution of outcomes and the quality of the process are of paramount societal importance.

[1] Delacroix S, 'Computer Systems Fit for the Legal Profession?' (2018) Legal Ethics, doi:10.1080/1460728x.2018.1551702

4 Defining Algorithms

The term algorithm can be understood in a variety of ways. A classic computer science definition, also echoed in many official reports, is ‘any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output’.[2] While this is a useful definition for computer scientists, the purpose of the algorithm is left in the background. Instead, we can think of an algorithm as part of the process of solving a computational problem. Given a specified computational problem which generally describes a desired input-output relationship, an algorithm describes a computational procedure which achieves this relationship.[3] Potential input-output relationships vary greatly: they could be transaction data and compliant tax returns, historical crime data and a schedule for police officer patrols, or an image of a person and the identity of that person.

Not all algorithms work in the same way, nor are all of them equally capable of estimating complex input-output relationships. In this report, we distinguish simpler, often rule-based algorithms from more complex algorithms.

Traditional rule-based systems have the **relationships between inputs and outputs crafted by hand**. These are often rule-based systems, like flowcharts, where the steps, methodologies and outcomes can be traced to pre-programmed instructions inputted by a human. They might be complicated, with hundreds or even thousands of steps, but generally represent pre-existing rules or theories.

More advanced algorithmic systems, which include machine learning approaches, are used with problems where pre-existing rules or theories do not capture the desired input-output relationships well. As a result, machines craft the relationship between inputs and outputs backwards from the data, usually without regard for human interpretability. In some cases, this can allow machines to make much more effective input-output connections – which computer scientists call predictions – than hand-crafted rule-based systems could. These systems are often referred to as non-parametric, as the parameters that define the input-output transformation are not defined in advance, but derived from the data or complex computational simulations.

One family of advanced algorithm has recently shown promise in several areas that were previously challenging for computational systems – machine learning.[4] It can be said that a machine ‘learns’ when, after being exposed to new data, it improves at a certain task according to the notion of performance we choose.[5] Machine learning techniques date back over a hundred years,[6] but from the 1960s onwards new techniques were created that allowed machines to construct more complex and interwoven input-output relationships.[7] Most recently, there has been a resurgence of interest in one particular decades-old form of machine learning, neural networks, after researchers demonstrated that with certain augmentations and on large datasets, it was especially effective on a range of difficult tasks.

[2] Cormen TH, Leiserson CE, Rivest RL and Stein S, *Introduction to Algorithms* (MIT Press 2009) 5.

[3] See the second definition presented in ibid, 5.

[4] Machine learning is by far not the only type of deployed algorithm of this type. For example, evolutionary algorithms or agent-based models are also used to understand and predict complex phenomena.

[5] Mitchell TM, *Machine learning* (McGraw Hill 1997).

[6] Linear regression, a commonly used statistical tool, is a form of machine learning, and is largely credited to Francis Galton. See generally Stanton JM, ‘Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors’ (2001) 9(3) Journal of Statistics Education DOI: 10/gd82dx.

[7] These include techniques such as support vector machines, random forests and neural networks.

There are three main forms of distinct machine learning tasks:

Supervised learning is where an algorithm is presented with a set of training data that contains labels of observations. The algorithm's learning is 'supervised' by these labels with the aim of establishing a generalisable input-output relationship.

Unsupervised learning is where there are no labelled observations or predictions, but the algorithm instead looks for structure, such as clusters, which can be interpreted later.

Reinforcement learning is where an algorithm is given input data, performs some action based on this data, and receives an outcome in response. This provides it with feedback it can use to improve its performance at the next action.

In addition to these three types, it is common to hear the term **deep learning** used. Deep learning is a type of neural network which can be used for the above tasks, called 'deep' due to the many layers of 'neurons' which data pass through on their way from input to output. Deep learning is a method of machine learning rather than a type of task or application, sometimes useful and sometimes not for the above tasks. Supervised, unsupervised and reinforcement learning can all be undertaken effectively using many machine learning approaches, of which deep learning is just one.

5 Introduction

Fundamental rights are core to a well-functioning democracy and are a condition for the ongoing trust and confidence of society in the machinery of the state and social wellbeing. The core purpose of this Commission is to ensure that fundamental rights and the rule of law continue to be respected during periods of wide-ranging social and technological change. The Commission is part of an effort to ensure that, if the nature of the justice system and the aspirations society associates with it are to change, this change is a conscious one, understood and meticulously stewarded, rather than a transformation beyond clear societal control and with unclear implications.

Algorithmic systems in the justice system are not new. Algorithms, such as risk scoring systems, have long been used by public agencies[8] and considered in regulation.[9] While some technological changes in the justice system, such as the digitisation of documents or law, seem unlikely to present significant challenges to fundamental rights, certain algorithmic systems have the potential to be significantly more value-laden and give rise to trade-offs between competing values. In determining how to make judgement calls between one value and another, we need rigorous processes and agreed principles.[10]

To help us understand these critical issues, we held four public evidentiary sessions, interviewed over 75 experts and read over 80 submissions of evidence.[11] These engagements spanned disciplines and sectors, with expertise concerning computing, regulation, political science, ethics, the rule of law, public policy, human rights and civil liberties, from both the public and private sectors.

This report centres on the criminal justice system, although it does touch on neighbouring areas of justice more broadly, and parallels from the recommendations contained within are likely to be useful across the domain.

5.1 Algorithms in Criminal Justice

Actors in the criminal justice sector, including police forces, crime labs, courts, lawyers and parole officers use algorithmic systems in a wide range of ways. Examples include:

- Photographic and video analysis, including facial recognition;[12]
- DNA profiling;[13]

[8] See generally Veale M and Brass I, ‘Administration by Algorithm? Public Management Meets Public Sector Machine Learning’ in Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (Oxford University Press 2019) doi:10/gfzvz8; Coglianese C and Lehr D, ‘Regulating by Robot: Administrative Decision Making in the Machine-Learning Era’ (2017) 105 Geo. L.J. 1147.

[9] They have also been the subject of statutory provisions in the UK. See for example the former Data Protection Act 1998, which contained exemptions for risk assessment systems applied to individuals for the purposes of crime prevention or the levying of taxes. Data Protection Act 1998 s 29(4)(a); a similar provision also exists in the current Data Protection Act 2018, sch 2 para 3(2).

[10] See generally Citron DK, ‘Technological Due Process’ (2008) 85 Wash U. L. Rev. 1249 (on public sector algorithmic systems and due process in the United States).

[11] See Annex 9.1.

[12] See *infra* section 7.2.

[13] Amankwaa AO and McCartney C, ‘The UK National DNA Database: Implementation of the Protection of Freedoms Act 2012’ (2018) 284 Forensic Science International 117.

- Individual risk assessment and prediction;[14]
- Predictive crime mapping;[15]
- Mobile phone data extraction tools;[16]
- Data mining and social media intelligence (SOCMINT).[17]

The variety of systems deployed today is examined below in section 7.

5.2 Drivers

Algorithmic systems appear to be increasingly deployed, and this report demonstrates that uptake and interest seem to be strongly expanding in a number of sectors. Several reasons can be posited as responsible for this growth.

5.2.1 Resourcing pressures

Across the public and private sectors, resource allocation and the need for greater efficiencies are driving the demand for greater use of automated and autonomous technologies.[18] This pressure is felt as much in the criminal justice field as in any other, especially with declining public funds year-on-year. Projections published in 2017 suggested the police service would be almost £200m in deficit by 2021.[19] This sits alongside general experiences and perceptions of austerity across the public sector: the percentage of people who are worried about whether the state will provide support for them in the years ahead has risen from 50% in 2010 to 70% in 2018, while the number of people who believe the government does too much fell from 64% in 2010 to 41% in 2018.[20] Funding for digital transformation is available, however, such as the Police Transformation Fund (PTF), totalling £42.7 million across 2018/19 and 2019/20 – creating strong incentives for forces to frame the development around digital technology to receive further central support.

5.2.2 Desires for increased proactivity

At the same time, the focus of UK criminal justice has been moving from the crime and the criminal to the victim and victimisation, with an emphasis on vulnerability placing greater demands on police forces to act in a generally anticipatory and preventative, rather than specifically reactive manner.[21] This focus on anticipating risk leads to predictive tools being seen as helpful in facilitating this aim in a cost-effective manner.

[14] See eg Moore R (ed), *A Compendium of Research and Analysis on the Offender Assessment System* (Ministry of Justice Analytical Series 2015) <https://perma.cc/W2FT-NFWZ>; Singh JP, Kroener DG, Wormith JS, Desmarais SL, and Hamilton Z (eds), *Handbook of Recidivism Risk/Needs Assessment Tools* (Wiley Blackwell 2018).

[15] Johnson SD, Birks DJ, McLaughlin L, Bowers KJ, and Pease K, 'Prospective Crime Mapping in Operational Context Final Report' (Home Office Online Report, Home Office 2007); Perry W, McInnis B, Price C, Smith S, and Hollywood J, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (RAND Corporation 2013).

[16] Privacy International, 'Digital Stop and Search' (Privacy International 2018).

[17] See generally Edwards L and Urquhart L, 'Privacy in Public Spaces: What Expectations of Privacy Do We Have in Social Media Intelligence?' (2016) 24 International Journal of Law and Information Technology 279.

[18] Alston P, 'Statement on Visit to the United Kingdom, by Professor Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights' (United Nations 2018); Dencik L, Hintz A, Redden J, and Warne H, 'Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services' (Data Justice Lab, Cardiff University 2018).

[19] National Audit Office, Efficiency in the Criminal Justice System (Comptroller General 2016).

[20] Deloitte and Reform, *The State of the State 2018-19* (2019).

[21] Williams E, Norman J, and Wunsch D, 'Too Little Too Late: Assessing Vulnerability' (2009) 3 Policing 355.

5.2.3 Data and infrastructure availability

The greater availability of linked datasets has also driven the use of algorithmic systems. Digitisation efforts have been ongoing in the public sector for many decades,[22] and accessing and combing data systems for an algorithmic tool no longer requires the heavy infrastructural investment it once did. Other technological infrastructures, such as the increased use of police GPS and flexible handheld or vehicle-based computers, also facilitates the greater use of algorithmic systems. Approaches to effectively evaluating and guiding police officers to hotspots, for example, rely on the development of GPS systems which can 'ping' each minute or more frequently, which are still not prevalent.[23]

5.2.4 New crime challenges

Technology is changing the nature of crime.[24] 'Cybercrime' describes two closely linked but distinct areas of criminal activity. Cyber-dependent crimes can only be committed technologically, with computing both the means and ends of a crime, such as ransomware or data destruction. Cyber-enabled crimes are traditional crimes which can be increased in scale or reach by the use of digital technologies, such as identity fraud or money laundering.[25] Digital money laundering, drone tracing, cybersecurity, marketplaces for illicit products on the 'dark web', harassment and abuse facilitated by connected technologies – all these are challenges where algorithmic responses are candidates for necessary and proportionate interventions to maintain the rule of law.[26] Algorithmic systems and artificial intelligence are thought to exacerbate challenges of cybercrime in a wide variety of ways.[27]

5.2.5 Access to justice

Both an all-too-often unmet need and a driver, the criminal justice is faced with an avalanche of problems, including a growing shortage of duty solicitors, increasing court closures, barriers to accessing legal aid, and crucial evidence not being available until the last minute.[28] Algorithmic systems might help bring efficiencies to the system through automation of rote tasks, or, more controversially, might take more value-laden decision support roles.[29]

Policy makers have increasingly focused their attention on how technology may be able to assist in bridging the need to access justice. In England and Wales specifically, HM Courts and Tribunals Service is undergoing a major programme of reform, predicated on extensive use of technology assisted justice – something which may herald a new era and broader context for the use of algorithmic systems in judicial processes and decision making.[30]

[22] Margetts H, *Information Technology in Government: Britain and America.* (Taylor and Francis 2012); Dunleavy P, Margetts H, Bastow S, and Tinkler J, *Digital Era Governance* (Oxford University Press 2006).

[23] Hutt O, Bowers K, Johnson S, and Davies T, 'Data and Evidence Challenges Facing Place-Based Policing' (2018) 41 Policing: An International Journal 339, 342–3.

[24] Europol, 'European Union Serious and Organised Crime Threat Assessment: Crime in the Age of Technology' (European Police Office 2017).

[25] HM Government, National.Cyber Security Strategy 2016 to 2021 (HM Government 2016).

[26] On the latter point, see generally Freed D, Palmer J, Minchala D, Levy K, Ristenpart T, and Dell N, 'A Stalker's Paradise: How Intimate Partner Abusers Exploit Technology' in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (ACM Press 2018).

[27] Brundage M and others, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (2018) https://maliciousaireport.com/.

[28] The Law Society of England and Wales*, Criminal justice system in crisis: Parliamentary briefing*. (The Law Society 2019). <https://www.lawsociety.org.uk/policy-campaigns/public-affairs/parliamentary-briefing/criminal-justice-system-in-crisis/>

[29] See generally HM Crown Prosecution Service Inspectorate and HM Inspectorate of Constabulary, *Delivering Justice in a Digital Age*. (Criminal Justice Joint Inspection 2016).

[30] See generally HM Government, The HMCTS reform programme (GOV.UK 2019). This is not a novel idea, but has been developing for decades. See further Susskind R, *Expert Systems in Law: A Jurisprudential Inquiry* (Clarendon Press 1989); Susskind R, *The Future of Law: Facing the Challenges of Information Technology* (Oxford University Press 1987).

Outside of criminal justice, online dispute resolution is a growing field. For example, Justice42 is an online platform in the Netherlands enabling people to work together to resolve their legal disputes, with help from experts if needed.

5.2.6 Belief in the efficacy of computational prediction

The last, overarching driver is the general increase in interest in data-driven systems in the public sector and beyond. While applying routinely captured data to better understand phenomena of public interest is not a new ambition or practice,[31] the notion that these datasets might contain insights that transform the nature of public service, particularly in light of newer applications of machine learning, has become.[32] Organisations in criminal justice do not wish to be left behind, and have mobilised behind recent discourse and interest in machine learning, artificial intelligence and data-driven decision making.

5.3 Benefits

Algorithmic systems have been promised to bring a range of benefits to different sections and functions. Below are some of the commonly held or cited benefits the Commission encountered in the context of criminal justice.

5.3.1 Efficiency from automation

One distinction we can draw between the benefits of algorithmic systems focuses on those which automate rote, relatively straightforward tasks where the outcome is not contentious, versus those which augment decisions to help organisations attempt to achieve outcomes they could not have without the insight from computational processes.[33] Automation of rote services, such as form-filling, checking, information retrieval and dissemination can bring strong benefits, assuming that the automation does not endanger the quality of tasks which cannot be effectively undertaken by machines alone, or remove important points of human contact and problem solving. Professor William Wong described this to the Commission as "[designing] a system whereby humans decide, and machines do the heavy lifting". Automation in public institutions has been a decades-long task, often promised to provide radical change but subject to organisational and political challenges along the way.[34] Machine learning technologies can help with automation, for example by helping automatically link systems which do not naturally work well together,[35] or by 'plugging in' types of information to processes that machines have previously had trouble parsing, such as handwriting or speech. Broadly, however, if automation is performing well on simple tasks without unexpected side effects, it has a limited impact on human and fundamental rights. However, these side effects can be complex, such as the impact of decreasing the availability of human interaction as an alternative to interacting with computers; accessibility

[31] Manzoni J, 'Big data in government: the challenges and opportunities' (*GOVUK*, February 2017) ⟨https://perma.cc/GF7B-5A2R⟩; Matthew Woollard, 'Administrative Data: Problems and Benefits. A perspective from the United Kingdom' in DuşA, Nelle D, Stock G and Wagner GG (eds), *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences* (SCIVERO Verlag 2014).

[32] See eg parliamentary interest in House of Common Science and Technology Select Committee, *Algorithms in Decision-Making*, (HC 351, Fourth Report of Session 2017–19, 2018); for a view elsewhere, see De Wetenschappelijke Raad voor het Regeringsbeleid (WRR) [Dutch Scientific Council for Government Policy], *Big Data in Een Vrije En Veilige Samenleving* [Big Data in a Free and Safe Society] (Amsterdam University Press 2016).

[33] Veale M and Brass I, 'Administration by Algorithm? Public Management Meets Public Sector Machine Learning' in K Yeung and M Lodge (eds), *Algorithmic Regulation* (Oxford University Press 2019) doi:10/gfzvz8.

[34] See e.g. Margetts H, 'The Automated State' (1995) 10 Public Policy and Administration 88; Margetts H, *Information Technology in Government: Britain and America.* (Taylor and Francis 1999).

[35] This is often described as *robotic process automation*, see generally Willcocks LP and Lacity M, *Service Automation Robots and the Future of Work* (SB Publishing 2016).

challenges for those on the other side of the digital divide or with impairments that make accessing online content difficult; or larger systemic impacts, such as changing the culture of litigation and justice through online courts and alternative dispute resolution systems. While automation can bring benefits, anything beyond very simple automation can have complex consequences.

5.3.2 Efficacy from augmentation

Algorithmic systems might not automate rote tasks, but they might provide new information and analysis that changes what the outcomes would have been under a traditional decision-making system. For example, certain challenges in the justice domain may be more effectively tackled using more granular targeting. Machine learning algorithms are designed to discriminate between cases, objects or individuals, and can tell them apart in ways that might be difficult in practice. Health and medical care have been promised a step change in treatment efficacy by moving to 'precision medicine', where interventions can be tailored based on the specific circumstances and characteristics of the patient. Such tailored interventions might hold promise in the justice sector too, for example in understanding how individuals can be provided with rehabilitative services best suited to their circumstances; delivering training courses while in detention; or identifying the leverage points in criminal networks most likely to disrupt their functioning. It is worth noting, however, that more granular interventions also come with the added risk of introducing discrimination inadvertently.

5.3.3 Auditability

Algorithms deployed in certain justice-related contexts may offer the ability to supply a greater degree of scrutiny of existing processes and outcomes.[36] Existing practices in complex areas often give practitioners a considerable degree of leeway in the grey areas where rules are less clear-cut.[37] This can be important for enabling a flexible system, but can also be a route for unwanted patterns, such as prejudice and discrimination. Where rules are implemented and enforced through a large number of practitioners, consistency of application can be jeopardised. Algorithmic systems – assuming their results are implemented faithfully and they are open for scrutiny – might offer a central point for examination and control of issues such as fairness.

5.3.4 Consistency and control

Relatedly to the questions of central auditability, consistency is seen as a clear virtue in procedural justice.[38] Rules are generally not seen as just if applied unequally to different populations and situations without appropriate reason. Appropriate algorithmic decision support to ensure a minimal level of consistency, as well as to counteract the behavioural biases of individual decision makers, might be beneficial in context where such variations are likely to be significant. With consistency comes considerations around the control of this consistency. The appropriateness of centralised control will always be reliant on a

[36] Tversky A and Kahneman D, 'Judgment under Uncertainty: Heuristics and Biases' (1974) 185 Science 1124.

[37] See generally Lipsky M, Street-Level Bureaucracy: Dilemmas of the Individual in Public Service (Russell Sage Foundation 2010).

[38] Leventhal GS, 'What Should Be Done with Equity Theory? New Approaches to the Study of Fairness in Social Relationships' in K Gergen, M Greenberg, and R Willis (eds), *Social Exchange: Advances in Theory and Research* (Plenum 1980); Colquitt JA, Conlon DE, Wesson MJ, Porter COLH, and Ng KY, 'Justice at the Millennium: A Meta-Analytic Review of 25 Years of Organizational Justice Research.' (2001) 86 Journal of Applied Psychology 425.

responsible mechanism for centralised decision making, but assuming such an approach exists, algorithmic systems might heighten, rather than limit, inconsistencies.[39]

5.3.5 Monitoring performance

The justice sector has historically been criticised for lacking reflection and uptake of evidence-based approaches, with police forces being accused of making decisions using other philosophies and processes, including “hunches and best guesses; traditions and habits; anecdotes and stories; emotions, feelings, whims, and stereotypes; political pressures or moral panics; opinions about best practices; or just the fad of the day”.[40] Algorithmic approaches can aid in reflection insofar as they mandate data collection easily amenable to analysis.

Furthermore, accuracy is not a single concept in an environment as complex as the justice sector. Just as different non-algorithmic interventions serve different populations differently, machine learning may force designers to make choices about different types of ‘performance’, such as between accuracy for particular subgroups, or between false positive and false negatives.[41] At times, some of these definitions will be at tension, and not all forms of performance of an algorithmic system are mathematically or statistically possible to achieve at once.[42] The decisions around how to measure and assess performance are often made implicitly in non-algorithmic systems: proper and rigorous use of machine learning forces responsible practitioners to make them more explicit, and may increase accountability and policy success as a result.

5.4 Dangers

There are many approaches to understanding dangers from algorithmic decision making in the justice sector, just as there are many ways of understanding general societal issues. One framing usefully separates instrumental concerns around the consequences of their use, misuse or errors; dignitary concerns which relate to the threat to individual human beings being respected as whole, free persons, and justificatory concerns, which surround the legitimacy or illegitimacy of a decisional system using algorithms.[43] To this framework we add the notion of systemic concerns, revolving around the changing nature of the criminal justice system and its interaction with society.

5.4.1 Instrumental concerns

Instrumental concerns surround the consequences of deploying algorithmic systems on individuals in the criminal justice system and in institutions tasked with ensuring justice is carried out fairly.

[39] To that effect, see the proposal that humans might wish to appeal to algorithmic decisions in Kamarinou D, Millard C and Singh J, ‘Machine Learning with Personal Data’ (2016) *Queen Mary School of Law Legal Studies Research Paper No. 247/2016*.

[40] Lum C and Koper CS, ‘Evidence-Based Policing’ in G Bruinsma and D Weisburd (eds), *Encyclopedia of Criminology and Criminal Justice* (Springer New York 2014).

[41] See generally Chouldechova A, ‘Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments’ (2017) 5 Big Data 153.

[42] Ibid.

[43] Further information on this distinction can be found in Kaminski ME, ‘Binary Governance: Lessons from the GDPR’s Approach to Algorithmic Accountability’ (2019) 92 Southern California Law Review __ doi: 10/gfzx54.

5.4.1.1 Bias and discrimination

Algorithmic systems encode assumptions and systematic patterns which can result in discriminatory outputs or downstream effects.[44] The way data used as input to systems is labelled, measured and classified is subjective and can be a source of bias. For example, taxonomies of reported crimes used by police forces or the demographics of individuals in the criminal justice system might not contain enough nuance, and this lack of nuance feeds forward into systems built on this data.

Training data itself is almost certain to be biased: there is no way to truly measure crimes committed in society, only proxies such as conviction, or more problematically, individuals arrested or charged.[45] If, as is commonly known, the justice system does under-serve certain populations or over-police others, these biases will be reflected in the data, meaning it will be a biased measurement of the phenomena of interest, such as criminal activity.

Furthermore, data might display societal patterns that we do not wish to reproduce or act upon without debate. For example, a certain postcode might have a higher prevalence of some type of crime, but it might not be the case that the area should receive additional policing as a result. If, hypothetically, an algorithm were used to shortlist new magistrates based on the existing population, it might underrepresent those from certain minorities or socioeconomic groups, and not connect to the current desires to diversify the pool of individuals chosen. Algorithms do this silently even when data such as ethnicity or socioeconomic status is not included as input data, as much information – such as education, address, or even more complex factors such as writing style – can act as proxies for information we might not want to base judgements on.

Biased and discriminatory systems might exacerbate themselves, particularly when those collecting data are the same as those acting on predictions. Areas more highly predicted become more heavily surveilled, and models become skewed without integrating data collected from a more evenly distributed collection.[46]

It should be noted that algorithmic systems need not be heavily biased in order to have a problematic effect. Even systems containing small biases are likely to result in cumulative disadvantage as their impact is compounded by the number of times and junctures where an individual or a community is impacted by one or more problematic systems.[47]

5.4.1.2 Oversimplification of complex issues

Algorithmic systems can only work with data that can be formalised and quantified. Even textual data or image data must be transformed into a form that can be mathematically manipulated. A core question for algorithms in the justice systems is what insight and information is lost in this process. Tacit knowledge, which can be difficult to formalise into rules, can be key for problem solving and understanding complex issues, but difficult to elicit

[44] See generally Barocas S and Selbst AD, ‘Big Data’s Disparate Impact’ (2016) 104 California Law Review 671.

[45] Veale M, Van Kleek M, and Binns R, ‘Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making’ in *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2018* (ACM Press 2018) doi:10/ct4s, 7.

[46] Ensign D, Friedler SA, Neville S, Scheidegger C, and Venkatasubramanian S, ‘Runaway Feedback Loops in Predictive Policing’ in Conference on Fairness, Accountability and Transparency (FAT* 2017) (PMLR 2018). On feedback loops in public sector machine learning more generally, see Veale M, Van Kleek M, and Binns R, ‘Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making’ in Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2018 (ACM Press 2018) doi:10/ct4s.

[47] Gandy Jr OH, ‘Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems’ (2010) 12 Ethics and Information Technology 29; Gandy Jr OH, *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage* (Routledge 2009).

for use algorithmically.[48] Relying on algorithmic systems might result in some decisions being made on a shallow view of evidence and without a deep, contextual consideration of the facts.

Furthermore, analytic capacity within public services is generally geared towards the short term, with long-term analysis neglected.[49] Algorithmic systems are not well suited to a longer view either, given that they are mostly able to produce outputs on immediately presented sets of data, rather than gathering intelligence about the future, exploring the dynamics of change and what future challenges might look like, and developing and testing responses to those challenges.[50] They may have roles within this process, but a singular focus on algorithmic systems may foster new forms of simplification and short-termism which do not adequately report responses to complex challenges in criminal justice.

Algorithmic systems, like all computing systems, are vulnerable to cyberattacks. Cyberattacks largely fall into three categories: those which compromise the confidentiality of information; the integrity of the decision-making process; or the availability of the system for use.[51] Use of algorithmic systems requires technical and organisational vigilance against actors who might wish to undermine the justice system.

Furthermore, while algorithmic systems might serve to streamline the justice system and increase economic efficiency, this raises the question of what redundant fallback measures exist when such systems fail, particularly if institutions and skillsets of humans without algorithmic support have weakened in the meantime.

Additionally, just as the potential for compromised electronic voting systems might serve to undermine trust in elections even if functioning well, the potential for algorithmic systems to pervert the course of justice might undermine trust in the justice system as a whole, particularly if salient compromises occur in other sectors or jurisdictions.

5.4.2 Dignitary concerns

Dignitary concerns focus on respect for individuals and communities in the criminal justice process.

5.4.2.1 Individuals not treated as such

Machine learning systems are similarity engines, seeking to find cases with traits that are similar to cases that were present in the training data and classify them similarly. Regardless of its accuracy, there are concerns that such a system "does not allow an individual to proclaim her individuality" and "violates her dignity and objectifies her as her traits, rather than treating her as a whole person".[52] Such concerns are compounded by how modern machine learning is able to work through correlation much more effectively than nascent efforts to automate the understanding of causation.[53] Membership of a group or similarity to other cases in a dataset do not cause criminality, victimisation, or other focuses of algorithms

[48] See generally Nonaka I, 'The Knowledge-Creating Company' (1991) Nov-Dec Harvard Business Review; in the context of public administration, see relatedly Lipsky M, *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service* (Russell Sage Foundation 2010).

[49] Parrado S, 'Analytical Capacity' in Martin Lodge and Kai Wegrich (eds), *The Problem-solving Capacity of the Modern State* (Oxford University Press 2014), 98.

[50] See generally Government Office for Science, *The Futures Toolkit* (HM Government 2017).

[51] This is known as the 'CIA triad'.

[52] Kaminski ME, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92 Southern California Law Review __ doi: 10/gfzx54.

[53] Pearl J and Mackenzie D, The Book of Why: The New Science of Cause and Effect (Allen Lane 2018).

in the justice system – but a heightened emphasis on correlation, simply because it is computationally possible, may cause the conflation of the two and place dignity at risk.

5.4.2.2 Dehumanised justice

Particularly in European countries, compared to the United States, there has been a regulatory and cultural aversion to systems being fully automated and void of a ‘human in the loop’.[54] The fear of ‘dehumanised justice’ – that a human might derogate responsibility to a decision-support system, or that a decision system was designed to not have human involvement from the start – has consequently resulted in provisions restricting such systems in UK and European law.[55] Particularly in relation to the coercive use of force by the state, the concern that a decision could be made without human involvement appears to be something that culturally, many citizens wish not to see happen.

The topic of automation bias has been a prevalent area of study, where humans may without due reason over- or under-rely on automated support.[56] In our evidence, this came up frequently. For example, Chief Constable Michael Barton of Durham Constabulary noted human decision makers may lack the confidence and knowledge to question or override an algorithmic recommendation.[57]

5.4.2.3 Loss of autonomy

There are concerns that algorithmic systems might manipulate people into situations they would not have been in otherwise. If an individual’s journey through the justice system is personalised in ways it may not have been otherwise – for example, being offered or denied rehabilitative services, or even just proactive information about these, based on algorithmic risk assessment – are those individuals being manipulated into paths which are less autonomous? Recent uses of machine learning inspired by behavioural economics to ‘nudge’ individuals into certain courses of action can be construed as problematic from a human rights perspective.[58] The idea that one is being constantly technologically surveilled with behaviour predicted may moderate an individual into a bland, constrained course of action, in fear of triggering systems designed to detect anomalies or deviation.[59] Relatedly, if proactive tools steer the actions of individuals, such as police officers, this might affect their ability to reflect upon their actions and develop the ‘moral muscle’ needed to support a fair and contemplative justice system.[60]

Additionally, when a decision is made about an individual, it is important to ask whether that decision could have been made otherwise, and discuss the acceptability of what would have been required in terms of autonomy. In machine learning, this has recently been described as ‘recourse’. Ensuring a system permits individuals actionable recourse (e.g. over factors

[54] Jones ML, ‘The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood’ (2017) 47 Soc Stud Sci 216.

[55] See eg, in the French context, Simon, MA, Communication de M. Alain Simon à la conférence annuelle des commissaires à la protection des données (Québec, septembre 1987), reported in Commission nationale de l’informatique et des liber- tés (CNIL), *8e Rapport au président de la République et au Parlement*, 1987 (La Documentation Francaise 1988) ⟨https://perma.cc/2NCW-R5Q3⟩ 243–248; see generally on the development of provisions around automated decision-making, Bygrave LA, ‘Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling’ (2001) 17 Computer Law & Security Review 17.

[56] See generally Skitka LJ, Mosier KL, and Burdick M, ‘Does Automation Bias Decision-Making?’ (1999) 51 International Journal of Human-Computer Studies 991.

[57] Oral evidence given by Chief Constable Michael Barton to the Commission.

[58] Yeung K, ‘“Hypernudge”: Big Data as a Mode of Regulation by Design’ (2017) 20 Information, Communication & Society 118.

[59] Cohen JE, “Examined Lives: Informational Privacy and the Subject as Object” (2000) 52 Stan. L. Rev. 1373, 1425–1426.

[60] Delacroix S and Veale M, ‘Smart Technologies and Our Sense of Self: Going Beyond Epistemic Counter-Profiling’ in K O’Hara and M Hildebrandt (eds), *Law and Life in the Era of Data-Driven Agency* (Edward Elgar 2019) doi:10/gfzvz9; Delacroix S, ‘Taking Turing by Surprise? Designing Digital Computers for Morally-Loaded Contexts’ [2018] arXiv:180304548.

that they are reasonably capable of adjusting) is key in contrast to non-actionable recourse, where a system would treat an individual differently only if they were to change immutable characteristics such as race or gender, or unreasonable-to-change characteristics such as marital status or offspring.[61]

5.4.2.4 Privacy

Privacy is a wide concept, and in many ways spans many types of concerns. Yet privacy can be endangered by algorithmic systems in many ways. Firstly, machine learning systems might explicitly be used to infer data or behaviours which are considered private to individuals or to groups from seemingly non-sensitive data.[62] Secondly, algorithmic systems might be used to retrieve information from contexts with expectations of privacy that were previously difficult to access at scale or at all, such as through police mobile phone extraction techniques, social media intelligence or facial recognition and 'smart' CCTV analysis.[63] Issues of privacy, which can be framed, assessed and measured in a variety of different ways, also touch upon many other concerns in this section.[64]

5.4.3 Justificatory concerns

Justificatory concerns relate to questions of legitimacy and procedural justice around decisions made and supported by algorithmic systems.

5.4.3.1 Opacity preventing scrutiny of justification

A common concern around algorithmic systems in society is their lack of transparency.[65] When an individual is faced with a decision or a measure in a criminal justice context, it is critical they can assess it was legitimate, justified, and ultimately, legal. Algorithmic systems are often proprietary in nature, and this can place barriers in the way of their availability for scrutiny both by the organisations deploying them and the individuals, communities and civil society organisations seeking to scrutinise them. Their technical details add an extra layer of opacity, as even if they are open, it may not be clear how they function. Open code has further limitations in a machine learning context, as it may not be understandable or auditable without access to the training data, which may be restricted for privacy or data protection reasons.[66] Furthermore, machine learning systems, in contrast to rule-based systems, are not designed with human interpretability in mind, but optimised instead for connection of input and output data with little regard for the comprehensibility of such connections.[67] Providing a trace of a system's 'thinking' is likely of little practical use in these situations to individuals attempting to assess whether an action was justified – something recognised

[61] Ustun B, Spangher A, and Liu Y, 'Actionable Recourse in Linear Classification' in *Proceedings of the ACM Conference on Fairness, Accountability and Transparency (ACM FAT*)* (ACM 2018).

[62] Edwards L and Veale M, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18, 32-38.

[63] Privacy International, 'Digital Stop and Search' (Privacy International 2018); Edwards L and Urquhart L, 'Privacy in Public Spaces: What Expectations of Privacy Do We Have in Social Media Intelligence?' (2016) 24 International Journal of Law and Information Technology 279.

[64] See generally O'Hara K, 'The Seven Veils of Privacy' (2016) 20 IEEE Internet Computing 86.

[65] See generally Edwards L and Veale M, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18; Pasquale F, The Black Box Society: The Secret Algorithms That Control Money and Information (Harvard University Press 2015).

[66] Note for example that the investigation into the recidivism system in the United States, COMPAS, by the journalistic organisation ProPublica, relied on a records release request of data on individuals facing parole decisions. This data would be unlikely to be released under freedom of information law in the United Kingdom due to data protection concerns; the United States has no comparable cross-sectoral privacy or data protection regime at the time of writing.

[67] Burrell J, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 205395171562251.

since the early days of legal expert systems.[68] While explanation facilities have been developed for many machine learning applications,[69] there are concerns that in many cases they may fail to help users grapple with the reality of complex systems, because they are unfaithful to their functioning, burdensome or overly challenging to understand, or because such systems use input data which humans cannot parse or understand the relevance of.[70]

5.4.3.2 Rule-making without scrutiny

When decision systems are introduced into public contexts such as criminal justice, it is important they are subject to the scrutiny expected in a democratic society. Algorithmic systems have been criticised on this front, as when developed in secretive circumstances or outsourced to private entities, they can be construed as rule making not subject to appropriate procedural safeguards or societal oversight.[71] Few provisions currently support civil society organisations or forms of collective oversight of algorithmic systems directly, leaving a significant accountability gap in need of remedy.[72] Where algorithms are deployed by private sector organisations directly, freedom of information law has limited current applicability.[73] Furthermore, it is unclear whether civil society organisations have the capacity to engage in meaningful oversight, particularly given the rapidity with which different systems are being deployed across the sector and across the world.[74]

5.4.3.3 Power and function creep from information infrastructures

Even where algorithmic systems and their associated informational infrastructures are deployed proportionally today, the tools deployed may not have appropriate safeguards to prevent them from being misused in the future. CCTV systems, for example, were deployed in an era where technology did not allow their contents to be analysed at scale automatically. The calculus underlying the appropriateness of these technologies may have changed with the advent of more advanced machine vision techniques. Just as there was strong opposition to the introduction of identity cards in the United Kingdom, algorithmic systems bring opportunities for powerful actors to engage in potentially illegal abuses of power, exacerbated by the opaque nature of the systems described above. For example, a considerable literature has emerged concerning how society might check that the algorithmic systems being deployed in practice are truly the ones that institutions claim to be deploying.[75]

5.4.4 Systemic concerns

Systemic concerns relate to the overall characteristics of the criminal justice system working in concert. In complex, value-laden systems, it is not always the case that ‘good’ parts make a ‘good’ whole. A systemic view is useful to identify how different components of a system

[68] Wick MR and Thompson WB, ‘Reconstructive Expert System Explanation’ (1992) 54 Artificial Intelligence 33.

[69] See generally Abdul A, Vermeulen J, Wang D, Lim BY, and Kankanhalli M, ‘Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda’ in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI ’18, ACM 2018).

[70] Edwards L and Veale M, ‘Slave to the Algorithm? Why a “Right to an Explanation” Is Probably Not the Remedy You Are Looking For’ (2017) 16 Duke Law & Technology Review 18.

[71] Citron DK, ‘Technological Due Process’ (2008) 85 Wash U L Rev 1249.

[72] Edwards L and Veale M, ‘Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?’ (2018) 16 IEEE Security & Privacy 46.

[73] Information Commissioner’s Office, Outsourcing Oversight? The Case for Reforming Access to Information Law (ICO 2019).

[74] Kemper J and Kolkman D, ‘Transparent to Whom? No Algorithmic Accountability without a Critical Audience’ (2018) Information, Communication & Society (2018) doi:10/gfdbp6.

[75] Some propose to use cryptographic tools to this end. See eg Kroll J, Huey J, Barocas S, Felten E, Reidenberg J, Robinson D, and Yu H, ‘Accountable Algorithms’ (2017) 165 University of Pennsylvania Law Review 633; Kilbertus N, Gascon A, Kusner M, Veale M, Gummadi KP, and Weller A, ‘Blind Justice: Fairness with Encrypted Sensitive Attributes’ in J Dy and A Krause (eds), *Proceedings of the 35th International Conference on Machine Learning*, vol 80 (Proceedings of Machine Learning Research, PMLR 2018).

might interact in unexpected ways. By definition, it is challenging to 'predict' systemic concerns, as they are often unexpected.[76]

5.4.4.1 Human rights

The use of algorithmic systems in the criminal justice system can engage a number of human rights. Determining which human rights are engaged, and how, is a complex exercise: our Human Rights framework precedes the advent of the technology in question, as do most of the legal frameworks used to justify the use of this technology.

Nonetheless, international human rights standards and norms should sit at the heart of their implementation, since international human rights law provides binding obligations on States, and their agencies.

A particularly important example of how algorithms may engage human rights involves the right to a fair trial in Article 6 of the European Convention on Human Rights (ECHR). At first glance, this fundamental human right might not seem in danger, as no stakeholders in the United Kingdom are discussing automating criminal justice decisions in the foreseeable future.[77] Discussions of electronic or online courts are broadly outside the scope of this criminal justice-focused report. However, the ECtHR has held that even when a single defect in procedural fairness would not violate the right to a fair trial, cumulative defects may do so.[78] Given the body of work emphasising the role of cumulative disadvantage in algorithmic justice,[79] this seems a significant concern.

Furthermore, while admissibility of evidence is generally not considered within the right to a fair trial as provided by the ECHR and instead a matter for UK law, clear UK consideration must be given to the admissibility of evidence obtained algorithmically, through techniques such as extraction of information from personal devices.

5.4.4.2 Changing nature of law

One way of looking at the law is as a codification of our societal values, which constantly evolve and change over time – be it laws around marriage, suffrage, rights or liabilities. One critical issue which emerged from evidence to this Commission was a concern that algorithmic systems in the justice sector which look at past data to predict the future run the risk of stagnation, holding the evolution of justice anchored in the past rather than free to evolve. In computing, this is known as 'concept drift', and is a challenge to understand and cope with in high-stakes environments.[80] Furthermore, when laws do change, practitioners work through principled analysis rather than simply from past incidents or cases. For the subset of algorithms that learn from past data, it is unclear what datasets those data-intensive systems could draw on so soon after a law has changed.

On this point, Barrister Jacob Turner told the Commission:

> "It is of course true that morals shift and ethics change but laws can be updated to fit new morals. When it became socially acceptable in wider

[76] See generally Perrow C, *Normal Accidents: Living with High Risk Technologies.* (Princeton University Press 2011).

[77] Sone have suggested and analysed the use of automated systems in assisting judges in less contentious, civil cases, however. See eg Hoogden RH van den, 'E-Justice, Beginselen van Behoorlijke Elektronische Rechtspraak' (PhD, Universiteit Utrecht 2007).

[78] *Mirilashvili v. Russia*, App no 6293/04 (ECtHR, 11 Dec 2008) para, 165.

[79] Gandy Jr OH, 'Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems' (2010) 12 Ethics and Information Technology 29; Gandy Jr OH, *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage* (Routledge 2009).

[80] Veale M, Van Kleek M, and Binns R, 'Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making' in *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2018* (ACM Press 2018) doi:10/ct4s, 7–8.

> society's eye for the laws on voting or marriage to change, the laws changed. We can use the same kind of democratic processes that were used to change the laws in any other area to do the same with regards to ethics for AI. With regards to the technical implementation of these, this is why we need to set-up the governance structures, including and involving experts who are able to track the changing laws and the changing rules and apply them with regards to the technology as it stands, at any given point, but this is an ongoing, dynamic process."

Compounding this, we can see systemic challenges emerging particularly in a common law context, given how law is made, and how law evolves and changes. As more predictive systems are used in criminal justice – potentially those which prevent difficult cases, or dispose of them in other ways – there could be a risk of inherent conservatism as a result. This would mean that:

> "…cases with a low success prediction are unlikely to be heard in court, in turn making organic changes within case law less likely. The latter changes indeed often depend upon an accumulation of previous, unsuccessful cases that trigger a growing number of dissenting voices (both within and without the judiciary). While there may be ways of developing tools that not only predict the chances of success in court, but also the likelihood that a particular case will eventually contribute to some organic evolution within case law, there will be little commercial incentives for the latter tools."[81]

Such inherent conservatism would change the criminal justice system in important and often difficult-to-perceive ways, and as such, it is important to take a systemic view on challenges concerning algorithmic systems in criminal justice.

.

[81] Delacroix S, 'Computer Systems Fit for the Legal Profession?' (2018) Legal Ethics, doi:10.1080/1460728x.2018.1551702.

6 Data protection and algorithms in criminal justice

UK data protection laws consist of an interwoven combination of the Data Protection Act 2018, the General Data Protection Regulation 2016, the Privacy and Electronic Communication Regulations (PECR), the fundamental right to data protection in the European Charter of Fundamental Rights, and Article 8 (right to respect for private and family life) of the European Convention on Human Rights. In the justice domain, the transposition of the Law Enforcement Directive 2018 into the Data Protection Act 2018 (Part 3) is especially important.

Data protection law applies whenever personal data is being processed. Personal data means:

> “any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person”.[82]

Personal and non-personal data are often interwoven. Benoit Van Asbroeck pointed out to the Commission that not only does “GDPR [give] a very large and dynamic definition of personal data”, but “even machine-generated data could in certain circumstances be considered personal data [...] likewise, the presence of only a minimal amount of personal data within a dataset ‘contaminates’ the set, rendering GDPR compliance of the entire dataset necessary”.

Some have argued that some complex machine learning models themselves might fall under the category of personal data.[83] The broad definition of personal data processing means that even if an algorithmic system is not personal data, querying it with individual records means that personal data is being processed, bringing the activity within the scope of data protection laws. Consequently, much algorithmic activity in the criminal justice sector will have data protection implications.

While law enforcement authorities have significantly more invasive powers relating to the collection and repurposing of personal data given the needs of the criminal justice sector, which stem from Part 3 of the Data Protection Act 2018, they are also subject to strict limitations meaning they can only be exercised in a narrow context, on the basis of legal powers, and in accordance with certain procedures.[84]

The Data Protection Act 2018 has four main roles:

- implementing derogations and specific national context for the General Data Protection Regulation 2016 (Part 2);
- transposing the Law Enforcement (Data Protection) Directive 2016 (Part 3);[85]

[82] GDPR, art 4(1).

[83] Veale M, Binns R, and Edwards L, ‘Algorithms that Remember: Model Inversion Attacks and Data Protection Law’ (2018) 376 Phil Trans R Soc A 20180083.

[84] Purtova N, ‘Between the GDPR and the Police Directive: Navigating through the Maze of Information Sharing in Public-Private Partnerships’ (2018) 8 International Data Privacy Law 52, 53.

[85] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA [2016] OJ L119/89 (Law Enforcement Directive).

- implementing the Council of Europe's data protection instrument, the modernised Convention 108, into data protection law in the area of the intelligence services, which are outside of the scope of EU law (Part 4); and
- specifying the regulatory powers of the Information Commissioner, other remedies and related frameworks (Parts 5-7).

Within the criminal justice context, processing falls into one of three regimes: the GDPR, the Law Enforcement Directive, or intelligence services processing. The distinction between the three is not always clear.

Processing falls into the Law Enforcement Directive (DPA 2018 Part 3) if it meets the personal scope conditions of being carried out by a 'competent authority', which is either a body described in Schedule 7 of the DPA 2018, or "any other person if and to the extent that the person has statutory functions for any of the law enforcement purposes", excluding intelligence agencies.[86] It must also meet the condition of material scope: the processing must be for the "purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security" – the 'law enforcement purposes' that also co-define the personal scope.[87] For example, if a police department was processing data about its employees' salaries, it would fall under the GDPR rather than DPA 2018 Part 3.

However, if a body that is not a competent authority within the meaning of section 30 of the Data Protection Act 2018 processes personal data for the purposes of preventing and detecting crime, such as a CCTV camera in a shopping centre, that processing may benefit from enabling provisions in the GDPR.

Compared to the Law Enforcement Directive, the GDPR has more substantive constraints, such as the scope of rights, but compared to the GDPR, the Law Enforcement Directive has stricter requirements of legality.[88] For example, while there is a default prohibition on the processing of special category data under the GDPR, which has to be overcome by establishing a basis for processing such as explicit consent of the data subject,[89] the Law Enforcement Directive does not prohibit such processing. It requires only that such processing be used "only where strictly necessary".[90] This necessity check is mainly implemented in UK law by requiring the controller to have an "appropriate policy document in place" and a condition identified in Schedule 8 of the DPA 2018.[91]

Regarding algorithmic systems, we can identify several concerns around the regime implemented by the Law Enforcement Directive, as well as the United Kingdom's transposition of the directive, which appears to omit areas that are important to this report.

[86] Data Protection Act 2018, s 30.

[87] Data Protection Act 2018, s 31.

[88] Purtova N, 'Between the GDPR and the Police Directive: Navigating through the Maze of Information Sharing in Public–Private Partnerships' (2018) 8 International Data Privacy Law 52, 60. See generally Information Commissioner's Office, *Guide to Law Enforcement Processing* (ICO 2019).

[89] GDPR, article 9(2)(a).

[90] Law Enforcement Directive, article 10.

[91] Data Protection Act s 35(5).

6.1 Automated decision making for law enforcement purposes

Data protection law prohibits decision with legal effect or (similarly) significant effect solely based on automated processing, unless certain conditions are met. Three main distinct regimes apply:[92]

1. GDPR Article 22*:* automated decisions not taken by a competent authority for a law enforcement purpose, justified by consent or necessity for a contract;[93]
2. Data Protection Act 2018 Part 2 (GDPR Article 22) [Part 2 ADM]: automated decisions not taken by a competent authority for a law enforcement purpose, but otherwise provided for in UK law;[94]
3. Data Protection Act 2018 Part 3 (Law Enforcement Directive) [Part 3 ADM]: automated decisions taken by a competent authority for a law enforcement purpose.[95]

Broadly, only cases 2 and 3 are relevant for our purposes, as limited (if any) criminal justice processing will occur on the basis of consent or necessity for contract. They will be distinguished as Part 2 ADM and Part 3 ADM.

The basic structure of these provisions is as follows:

A ‘significant decision’ based solely on automated processing must be required or authorised by law, else it is prohibited. The definition of ‘significant decision’ depends on whether it falls within the scope of Part 2 or 3 of the DPA 2018. It either:

1. produces a legal effect concerning the data subject or similarly significantly affects them (Part 2 ADM); or
2. produces an adverse legal effect concerning the data subject or significantly affects them, and is made by a competent authority for law enforcement purposes (Part 3 ADM).

Where such a decision is authorised or required by law, the controller must notify the data subject in writing as soon as reasonably practicable, and, upon the request of the data subject, within one month of receipt of the notification, reconsider the decision or take a new decision not based solely on automated processing.

The Commission has several concerns around the protection provided by these provisions, and broadly recommends strengthening algorithmic protections in data protection. This overarching recommendation is broken down further in the sections that follow.

Recommendation 2 Strengthening Algorithmic Protections in Data Protection – The protections concerning algorithmic systems in Part 3 of the Data Protection Act 2018 should be clarified and strengthened.

6.1.1 Risk of rubber-stamping

The definition of solely based on automated processing is unclear in UK law. For example, if a police officer is involved in executing a decision, such as a stop-and-search choice, or a visit to an individual identified as vulnerable, does that decision cease to become solely

[92] A fourth regime relating to intelligence service processing can be found in the Data Protection Act 2018 s 97–98, based on the provisions of the Council of Europe Convention 108+, however in practice this is unlikely to be of utility as it will be excluded by s 110-111 on the grounds of national security.

[93] GDPR art 22.

[94] Data Protection Act 2018, s 14.

[95] Data Protection Act 2018, s 49-50.

automated, and therefore fall outside of the provisions? Few decisions by police forces are executed by machine, and a human is involved in turning intelligence into action.

Liberty gave evidence to the Commission of "the flawed notion of a 'human in the loop'". Their view was that "[w]hile the idea of having human involvement or oversight of an algorithmic decision-making process may sound reassuring, there is a lack of evidence as to our ability as humans to provide meaningful intervention over algorithms and decisions made by machines". This view was also held by Silkie Carlo from Big Brother Watch, who suggested to the Commission that two important amendments are required to the Data Protection Act 2018. First, decisions that engage individuals' human rights must never be purely automated decisions; second, automated decisions should be more clearly defined as those lacking meaningful human input.

Before the Data Protection Bill (as it then was) was laid before Parliament, European Data Protection Board (EDPB) – the collective group of EU data protection regulators – had already stated that human input must be meaningful, and individuals must have the authority and competence to challenge the decision.[96] They note that "[t]o qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the relevant data". These are welcome clarifications, albeit ones which the UK Government declined to place explicitly in statute during the passage of the Data Protection Act 2018. If the United Kingdom leaves the European Union, the status of EDPB guidance will be unclear.

In comparison, French administrative law has automated decision provisions which apply to any time when individuals are subject to an 'algorithmic treatment', rather than just to fully automated decisions.[97] Furthermore, the Dutch Scientific Council for Government Policy have recommended that decision that are 'semi-automated' be within scope of data protection law.[98]

Sub-Recommendation 2.4 Meaningful Human Intervention – The Data Protection Act 2018 should be amended to specify the nature of the input needed to not to be a decision "based solely on automated processing" and trigger Article 22 of the GDPR, Section 14 and 49 of the Data Protection Act 2018.

6.1.2 Omission of discrimination provisions in DPA 2018 Part 3

Both the GDPR and the Law Enforcement Directive have clear provisions around discriminatory profiling and/or automated decision-making which problematically are not replicated in UK law.

[96] Article 29 Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679* (WP251rev.01 2018); see further Veale M and Edwards L, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling' (2018) 34 Computer Law & Security Review 398. At the time, the EDPB did not exist, and was formed under the Article 29 Data Protection Working Party. (A29WP) specified in the Data Protection Directive 1995. The GDPR created the EDPB as a new official European body, and it adopted much of the previous A29WP's guidance in relation to the GDPR, and so the new name is used even for these past documents.

[97] Veale M and Edwards L, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling' (2018) 34 Computer Law & Security Review 398.

[98] De Wetenschappelijke Raad voor het Regeringsbeleid (WRR) [Dutch Scientific Council for Government Policy], Big Data in Een Vrije En Veilige Samenleving [Big Data in a Free and Safe Society] (Amsterdam University Press 2016).

For Part 2 ADM, Recital 71 of the GDPR specifies that controllers should ensure that their profiling systems do not result in discriminatory effects on the basis of special categories of data. It states that:

> "In order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures […] that prevent, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect."

The United Kingdom will import recitals into UK law as part of the European Union (Withdrawal) Act 2018 if it leaves the European Union, and allow them to be used in 'casting light on the interpretation to be given to a legal rule' as they were previously.[99] However, the Commission is concerned in relation to Part 2 ADM that a recital, as it does not have the status of a legal rule, is not protective enough in this important domain.

Part 3 ADM should have replicated the stronger, more specific and binding provisions on discrimination that are found in the Law Enforcement Directive. These state that:

> "Profiling that results in discrimination against natural persons on the basis of special categories of personal data referred to in Article 10 shall be prohibited, in accordance with Union law."[100]

These provisions do not simply apply to automated decision making, but to profiling more generally – including its application in non-solely automated settings. As described in the DPA 2018 Part 3, profiling means:

> "…any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to an individual, in particular to analyse or predict aspects concerning that individual's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements."[101]

The UK Government may, in its reasoning to not implement Recital 71 or the Law Enforcement Directive's provisions on discrimination, point to the protections under UK equality law. These are insufficient in algorithmic domains, because they need a careful consideration of how enforcement issues play out. The Information Commissioner's Office has a programme on automated decision making, considering issues such as fairness and transparency, and splintering enforcement over different domains is problematic.[102]

Sub-Recommendation 2.6 Discrimination Provisions in Data Protection – The Government should explicitly transpose Article 11(3) of the Law Enforcement Directive concerning the prohibition on discrimination of algorithmic systems, and make explicit statutory provisions for ensuring that Part 2 ADM are not discriminatory under the powers to derogate from the GDPR provided by Article 22(2)(b), GDPR.

[99] Explanatory Notes to the European Union (Withdrawal) Act 2018, footnote 27/

[100] Law Enforcement Directive, art 11(3).

[101] Data Protection Act, s 33(4).

[102] The ICO's AI Auditing Framework and the ExplAIn Project are examples of the ICO work in this area.

6.1.3 Limited protection of groups

While, as discussed in section 6.1.2, it would be desirable to have stronger data protection provisions around discrimination, there are also limitations to this approach. Data protection as a regime has highly individual foundations, and this has been reflected in its provisions. Many problems in algorithmic bias can best be understood as disadvantaging a group or a community, rather than an individual.[103] It may also be affecting individuals and groups cumulatively over time, rather than being 'significant' at a particular point.[104]

As a result, it becomes important to consider whether a use of an algorithm will affect a group rather than an individual. This formation already exists in UK law, such as in the public sector equality duty. This duty states that while exercising its functions, a public authority must, *inter alia*, have due regard to the need to "advance equality of opportunity between persons who share a relevant protected characteristic and persons who do not share it".[105] Having due regard to this includes having due regard to the need to:

- remove or minimise disadvantages suffered by persons who share a relevant protected characteristic that are connected to that characteristic;
- take steps to meet the needs of persons who share a relevant protected characteristic that are different from the needs of persons who do not share it;
- encourage persons who share a relevant protected characteristic to participate in public life or in any other activity in which participation by such persons is disproportionately low.[106]

These all have notions of the group which might be discriminated against, whereas data protection law requires demonstrating a 'significant' effect on an individual. Even proving this may be tricky and exclusionary, whereas it is easier for civil society organisations, for instance, to amass evidence about the mistreatment of groups. However, the public sector equality duty is primarily exercisable only at considerable cost through the mechanism of judicial review. Introducing these notions into data protection law would be useful in making it clear what a 'significant' decision might look like.

Sub-Recommendation 2.7 Significant Decisions and Groups – The Data Protection Act 2018 should be amended to clarify that a decision can be considered 'significant' if it impacts upon a protected or otherwise salient group to which a natural person belongs, rather than considering only impacts upon a single individual.

6.1.4 Omissions of transparency provisions in DPA 2018 Part 3

The Law Enforcement Directive and its transposition omit the provisions on transparency of automated decisions present in the rest of data protection law. The GDPR maintains that, for decisions that qualify for Article 22 protection, the individual should be provided, both before processing begins and upon request, with "meaningful information about the logic of processing".[107] The Law Enforcement Directive has no such provisions. Furthermore, the

[103] Edwards L and Veale M, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18.

[104] Gandy Jr OH, Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage (Routledge 2009); Gandy Jr OH, 'Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems' (2010) 12 Ethics and Information Technology 29.

[105] Equality Act 2010, s 149(1)(b).

[106] Equality Act 2010, s 149(3).

[107] See generally Selbst AD and Powles J, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233.

Law Enforcement Directive allows for competent authorities to refuse to provide information requested under the right of access, and even to provide a ‘neutral’ reply – to neither confirm nor deny the possession of this information.

This is an important provision, and it is clear that many information releases from competent authority might undermine the justice process. Yet the carve-outs prevent that from happening, and in addition, data subjects have the option of going to the Information Commissioner and asking her to request the data subjects’ data on their behalf and check the legality of the processing.[108] The Commission feels that by default, the ‘meaningful information’ provisions should be provided to individuals, unless the competent authority has reason not to release it, in which case the logic can be assessed by the Information Commissioner. Without that pathway, assessing the legality of algorithms is difficult in practice. Furthermore, it is crucial that there remain incentives for explainable and interpretable systems, and the requirement of these provisions in Part 3 as well as in Part 2 and the GDPR provide important incentives in this regard.

Sub-Recommendation 2.1 Transparency Rights – The transparency provisions concerning profiling and algorithmic decision-making in the GDPR (particularly Articles 13(2)(f), 14(2)(g), 15(1)(h)) should be mirrored for law enforcement in Part 3 of the Data Protection Act 2018 (s 44–45) and subject to the same balancing test for disclosure, rather than removed entirely.

6.2 Unclear provisions around public-private partnerships

Many algorithmic systems are developed in, or in close collaboration with, the private sector. This has been highlighted as problematic, and a barrier to challenging algorithms in courts.[109] From a data protection standpoint, it is important to identify the ‘controller’ of the processing being undertaken. A controller is an entity that determines the ‘means and purposes of processing’. At times, this might be the designer of an algorithmic system, such as an upstream contractor – potentially one that has never even seen personal data.[110]

Where this happens, it is unclear how joint liability and controllership might fall legally. Scholars have indicated that there may be a case where it seems appropriate to have a joint controllership between a competent authority and a ‘normal’ entity, and it is unclear whether the GDPR, the Law Enforcement Directive, or even both apply.[111]

Sub-Recommendation 2.5 Public Private Partnerships – The ICO should provide guidance on how the Data Protection Act Part 3 functions in the contexts of public-private partnerships and algorithmic systems.

[108] Data Protection Act 2018, s 51.

[109] See, in a US context, AI Now Institute, Litigating Algorithms: Challenging Government use of Algorithmic Decision Systems (New York University 2018); Pasquale F, The Black Box Society: The Secret Algorithms That Control Money and Information (Harvard University Press 2015).

[110] The CJEU has held that no personal data needs be seen for an entity to be considered a data controller. See Case C210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH* ECLI:EU:C:2018:388 para 38.

[111] See generally Purtova N, ‘Between the GDPR and the Police Directive: Navigating through the Maze of Information Sharing in Public–Private Partnerships’ (2018) 8 International Data Privacy Law 52.

6.3 Data Protection Impact Assessments

Data Protection Impact Assessments, which assess the impact of processing that is potentially high risk on the rights and freedoms of individuals, are required to different degrees in the GDPR and Part 3 of the Data Protection Act 2018.

Data Protection Impact Assessments have been characterised as ‘meta-regulation’,[112] making organisations responsible for their own efforts to self-regulate, and facilitating the more holistic adherence with a regulation which can be technologically and socially complex, and heavily value-laden.

DPIAs do not need to be published by default: no provision requires their publishing either in Part 3 of the Data Protection Act 2018 or the GDPR. However, when used in criminal justice, it does appear to be proportionate to publish DPIAs, or to make a public-facing version. Such text will often already be available upon request by using the Freedom of Information Act 2000, but it does require an empowered data subject to be aware of the existence of this processing (or updates to any DPIA, which should take place on a continuous basis). This seems unnecessary, and not displaying the openness that would or should be expected in criminal justice.

Sub-Recommendation 2.3 Data Protection Impact Assessments – Where Freedom of Information tests restrict release, a bespoke public-facing version of a data protection impact assessment concerning a consequential algorithmic system in criminal justice should be proactively published.

6.4 Further issues

As has been noted, many data protection issues in criminal justice bring significant possible carve-outs that the UK or the Secretary of State can make. Despite these, it is important to note that such carve-outs will need to be in line with the fundamental rights to privacy and data protection. If the UK leaves the EU, the fundamental right to data protection will no longer directly apply. Although privacy and data protection are entwined, there are also distinctions between how they are both operationalised in the CJEU, and how the ECtHR operationalises privacy and refers to data protection.[113] At this sensitive time of technological and societal change, it is important that any change in the UK fundamental rights regime does not result in a slippage of real protections for vulnerable individuals in a justice context.

One area that is often under-emphasised in a justice context is the privacy of employees within the justice system. There is a general trend which some consider worrying towards increased employee surveillance,[114] particularly within the quantification and ‘target culture’ that has emerged as a result of New Public Management.[115] Algorithmic systems for performance management of individuals in the justice system, if sufficiently invasive, might

[112] Binns R, ‘Data protection impact assessments: A meta-regulatory approach’ (2017) 7(1) International Data Privacy Law 22.
[113] Kokott J and Sobotta C, ‘The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR’ (2013) 3 International Data Privacy Law 222.
[114] For a UK and EU context, see Edwards L, Martin L, and Henderson T, ‘Employee Surveillance: The Road to Surveillance is Paved with Good Intentions’ (SSRN Scholarly Paper, Social Science Research Network 18 August 2018); in a US context, see Ajunwa I, Crawford K, and Schultz J, ‘Limitless Worker Surveillance’ (2017) 105 Calif L Rev 735.
[115] Hood C, ‘Public Service Management by Numbers: Why Does It Vary? Where Has It Come From? What Are the Gaps and the Puzzles?’ (2007) 27 Public Money and Management 95.

cause ethical and legal issues in relation to those individuals. If certain aspects of justice do become more remote and not face-to-face, there is often a tendency for managers of these workers to attempt to exercise control to avoid the exercise of discretion, which is of limited efficacy and may encourage individuals to hide their actions or exercise necessary discretion in other ways.[116] The ECtHR has held on several occasions that monitoring in the workplace can, in certain conditions, be a violation of the right to privacy.[117]

7 Algorithms in Criminal Justice Today

Algorithms are currently used by a variety of public bodies operating across the justice system. Through the evidence submitted to the Commission, it is clear there is widespread and growing use of algorithmic tools, and that trend looks to be an upward one. The uses of these technologies in the criminal justice system gives rise to, at its most basic, debates about the trade-offs between the potential benefits discussed (see *supra* section 5.3), such as increased crime prevention and detection, and the dangers (see *supra* section 5.4), such as risks to civil liberties and fundamental human rights.

This report does not aim to present a comprehensive review of these technologies as they are deployed today. Instead, the section that follows presents an indicative framing and list of algorithmic technologies encountered by the Commission, with particular focus on those deployed in England and Wales.

No central information exists on the number of forces using algorithmic tools, the crimes or other issues these technologies are applied to or the stage of deployment they are at. A freedom of information based study in the United Kingdom in 2016 asked all police forces whether they used any sort of computational or algorithmic data analysis or decision making in relation to the analysis of intelligence, and to confirm the nature and purpose of any such algorithms, and 14% of police forces reported affirmatively.[118] The civil society organisation Liberty similarly made 90 freedom of information requests to police forces in 2018, with 14 returned affirmatively.[119]

7.1 Predictive Hotspot Policing

Much attention around the fairness and accountability of algorithmic systems has centred on the use of them for determining the timing and location of police interventions. These technologies entered the public consciousness largely through reporting on deployments of the American software product PredPol. In the United Kingdom however, deployments of very similar technology have a much longer history. ProMap, built by researchers at the Jill Dando Institute of Crime Science, University College London around 2004,[120] was deployed and evaluated by the Home Office in the East Midlands in 2005/6.[121] This system for data-driven prospective or predictive analysis of burglaries builds on the history in UK police

[116] Jorna F and Wagenaar P, ‘The “Iron Cage” Strengthened? Discretion and Digital Discipline’ (2007) 85 Public Administration 189.
[117] See generally ECtHR, *Surveillance at workplace* (Council of Europe, Press Unit, November 2018).
[118] Oswald M and Grace J, ‘Intelligence, Policing and the Use of Algorithmic Analysis: A Freedom of Information-Based Study’ (2016) 1 Journal of Information Rights, Policy and Practice.
[119] Couchman H, ‘Policing by Machine’ (Liberty 2018).
[120] Bowers KJ, Johnson SD and Pease K, ‘Prospective Hot-Spotting: The Future of Crime Mapping?’ (2004) 44 British Journal of Criminology 641.
[121] Johnson SD, Birks DJ, McLaughlin L, Bowers KJ, and Pease K, *Prospective Crime Mapping in Operational Context: Final Report* (Home Office Online Report 19/07 2007).

forces of using geographic information systems (GIS) approaches for descriptive hotspot analysis. In these traditional systems supporting hotspot policing, maps were built manually by analysts based on data such as recorded crimes and phone calls made by the public.[122]

Predictive hotspot tools make statistical forecasts about where future crime may take place or where future police interventions may have a positive impact. The theoretical basis of much predictive policing research centres on theories seeking to explain the empirical phenomenon of repeat victimisation – that is, that recent victims of crime are temporarily at higher risk of crime than non-victims. Two main explanations have been proposed: that repeat victimisation is the result of returning offenders, or that it is a result of vulnerable individuals being 'flagged' as easier targets.[123] Theoretically, this was expanded to include the idea that crime might 'spread' to neighbouring locations,[124] and that 'dosages' of pre-emptive police patrols in these areas might act as a deterrent.[125]

Some other predictive hotspot tools seek to go beyond theory-based interventions, and integrate data from other sources which does not have a well-understood connection to crime, such as the weather, or sociodemographic features of the geographic area where crime being predicted.[126] These tools in particular raise heightened concerns around bias and discrimination through the use of proxy variables (see *supra* section 5.4.1.1), as they have discarded the concept of having a theoretically grounded model in favour of a 'black box' where any structured input data might serve to increase the accuracy of the predictions.

West Midlands Police uses a software with these features called MapInfo, which can correlate when crimes occur, the seasonality, the days of the week and times of the day, as well as crime data and antisocial behaviour reports. MapInfo is a general-purpose GIS tool first introduced in 1986, used by a range of sectors including transport, mining, insurance and telecommunications for descriptive and predictive purposes. In West Midlands, it is being used by around 150 trained staff and officers who employ it for varied crime mapping purposes, but it is unclear how far MapInfo is being used predictively or in line with more traditional hotspot-building technologies.[127]

Some police forces have built models and algorithmic systems in-house using commercially available platforms. **Avon and Somerset Police** have built such models using a platform procured from IBM called SPSS Modeller, which provides a visual programming language for machine learning tool creation. They additionally use a dashboarding tool for their predictive systems, QlikSense.[128] Avon and Somerset have not released a public-facing report on their modelling efforts, such as the variables their models contain, or monitoring and evaluation reports from any pilots undertaken. They did, however, co-fund a PhD position together with the University of the West of England Faculty of Health and Social Sciences on 'The Impact of Predictive Analytics on Policing Practice and Effectiveness', which commenced in 2015[129] – however, no work from this initiative has yet been published.

[122] College of Policing, *The Effects of Hot-Spot Policing on Crime: What Works Briefing* (College of Policing, September 2013).
[123] Pease K, Repeat Victimisation: Taking Stock (Home Office 1998).
[124] Bowers KJ, Johnson SD and Pease K, 'Prospective Hot-Spotting: The Future of Crime Mapping?' (2004) 44 British Journal of Criminology 641.
[125] Johnson SD, Birks DJ, McLaughlin L, Bowers KJ, and Pease K, *Prospective Crime Mapping in Operational Context: Final Report* (Home Office Online Report 19/07 2007).
[126] See e.g. Azavea, *HunchLab: Under The Hood* (Azavea 2015).
[127] West Midlands Police, Freedom of Information Request 07/11/2016 10951_16.
[128] Dencik L, Hintz A, Redden J, and Warne H, 'Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services' (Data Justice Lab, Cardiff University 2018).
[129] Centre for Legal Research, *Annual Report 2013/2014* (University of West of England 2014), 6.

Kent Police utilised the PredPol system for five years up until 31 March 2018, and are currently working to develop a similar programme internally. Unlike other police forces, some evaluation data is available from Kent Police's use of PredPol in the form of an academic study where Kent Police's existing hotspot prediction practices were compared to the outputs of the PredPol system.[130]

Academic collaborations are common between UK police forces and universities around statistical systems. **London Metropolitan Police** partnered with University College London on a £1.4m Research Councils UK-funded project concerning geospatial predictive modelling for policing between 2012 and 2016.[131] Among other activities, this project sought to understand whether the grid-based predictions used in technologies such as ProMap and PredPol could be superseded by one that focussed on predicting crimes using street networks.[132] Whether the force use this technology currently in any capacity is unclear. **West Yorkshire Police** are also currently working with University College London on a system to predict areas at high risk of crime, which is currently in a pilot stage. The proposed system will suggest a 'patrol plan' for staff using the system. **Norfolk Police** reported to Liberty that they are currently developing an algorithmic tool to assist in deciding whether burglaries require further investigation alongside the University of Cambridge. No tests or pilots of the tool were ongoing at the time of the request.[133]

7.1.1 Concerns

There are a range of concerns surrounding hotspot policing in general that have become salient in recent years.

The first surrounds the feedback effects that follow from having those that collect data for a system be the same as those who follow the system's guidance. Simulations indicate that if the data is only sourced this way, then systems can fall into feedback loops, where areas that are overpoliced become more so, to the detriment of other locations.[134] Where there are geographic divisions with strong demographic or socioeconomic characteristics, there are concerns that this could fuel discriminatory policing.

These feedback effects exacerbate a general issue of bias inherent to many risk mapping systems. An inherent tension exists between the need for predictive mapping to be based on recent, timely data, and the observation that police can only rapidly observe whether crimes have been reported or individuals arrested or charged – not whether a conviction eventually occurs. Reporting of crime might suffer from biases, such as the propensity for individuals of one culture to report those of another to the police for e.g. noise disturbance; something which may be more illustrative of community tensions or misunderstandings rather than of criminal or antisocial activity.[135] Arrests and convictions suffer from a multitude of potential biases, ranging from differential propensity to be stopped and searched to likelihood of conviction based on background, exacerbated by the dwindling availability of legal aid.

[130] Mohler GO, Short MB, Malinowski S, Johnson M, Tita GE, Bertozzi AL, and Brantingham PJ, 'Randomized Controlled Field Trials of Predictive Policing' (2015) 110 Journal of the American Statistical Association 1399.

[131] See https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/J004197/1.

[132] Rosser G, Davies T, Bowers KJ, Johnson SD, and Cheng T, 'Predictive Crime Mapping: Arbitrary Grids or Street Networks?' (2017) 33 J Quant Criminol 569.

[133] Norfolk Police, Response to Freedom of Information request by Liberty (July 2018).

[134] See generally Ensign D, Friedler SA, Neville S, Scheidegger C, and Venkatasubramanian S, 'Runaway Feedback Loops in Predictive Policing' in *Conference on Fairness, Accountability and Transparency (FAT* 2017)* (PMLR 2018).

[135] Veale M, Van Kleek M, and Binns R, 'Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making' in *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2018* (ACM Press 2018) doi:10/ct4s.

Opacity of deployed algorithmic systems is a constant concern in the policing context. The United Kingdom does have a degree of general transparency provided through the police data portal, police.uk, where datasets around crime prevalence and arrest throughout the UK can be browsed and downloaded. Models deployed by police however are not available either on this site or on public sector open data portals. Campaigners have pointed to general cultures of opacity; however, it is likely that other issues are also at play here, such as vendor restrictions and software licences, as well as a general lack of in-house capacity for providing scrutiny capacity as well as creating and maintaining models. Open data practices do not come without significant investment, and given the resourcing pressures on police, considering that they are adequately funded to enable scrutiny of algorithmic systems is likely an important consideration for central government.

There are concerns that over-reliance on predictive mapping systems might result in officers deferring judgement over their patrols to the algorithmic system, which could come with implications for the quality of policing in cases where the integration of difficult-to-quantify knowledge is key. Automation bias, as it is known in psychological and human-computer interaction research, can occur both ways: individuals can both over- and under-rely on computer systems.[136] While it might be assumed that police officers would be likely to over-rely on systems, in practice, evidence is mixed, with officers in some roles appearing anecdotally to follow decision-support systems closely, and officers in other roles expressing strong dislike of them.[137] Furthermore, while some police forces report in interview research maintaining a strong role for traditional analysts making predictive maps in augmenting them with qualitative data that is hard to encode in a theory-based model like repeat offending, such as whether a burglar has been caught or whether the locks in a public building have been changed, there is reported organisational concern that such routines and practices might diminish over time, or when forces inherit technologies rather than develop them in-house and learn the lessons ‘the hard way’.[138]

Broadly, these concerns culminate in worries that algorithms such as hotspot policing, which may not use personal data at the point of being queried (though may do so in training), fall outside of many of the protections that regimes such as data protection provide. Police have broad remit to patrol in a way that they consider effective, but these decision-support tools bring reason for societal concern in complex environments where discrimination is a real concern and technical capacity may not be sufficiently high to deal with some of the difficult issues researchers are highlighting. We return to such concerns in our recommendations in the concluding section of this report.

7.2 Facial Recognition in Policing

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

[136] See generally Skitka LJ, Mosier KL, and Burdick M, ‘Does Automation Bias Decision-Making?’ (1999) 51 International Journal of Human-Computer Studies 991.

[137] Veale M, Van Kleek M, and Binns R, ‘Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making’ in *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2018* (ACM Press 2018) doi:10/ct4s.

[138] Ibid.

to both more efficiently enable the identification of some individuals, as well as with a view to identify some sought-after individuals who probably would not have become known to police within those contexts otherwise.[139]

Only the first of these applications is in mainstream use today in the justice system, and so will be focused on below.[140]

As of 15 July 2016 there were 16,644,143 images enrolled in the facial image recognition gallery of the Police National Database, searchable using automated facial recognition software.[141] As of that date, the Home Office did not record how many of those faces belonged to individuals not charged of any offence.[142] As of January 2018, that number is reported to have dropped to 12.5m for reasons unknown.[143]

In England and Wales, the highest profile uses of facial recognition technologies in public spaces by the police consist of trials run by the **London Metropolitan Police**, the **South Wales Police** and **Leicestershire Police**. All three forces are trialling technologies produced by NEC, a Japanese firm.[144] The forces have limited ability to oversee or alter the software provided by NEC without the firm taking initiative,[145] with the deputy chief constable of South Wales Police noting that "the tech is given to [them] as a sealed box… [South Wales Police] have no input – whatever it does, it does what it does".[146] This software can run in two main modes: an Identify mode, which functions on pre-recorded images and compares them to a database of held images, which can be in the hundreds-of-thousands, or a Locate mode which works on live-streamed video but against a smaller, pre-filtered database of a few hundreds or thousands.[147] Both approaches provide a list of likely individuals to a staff member who manually examines the results.[148]

Leicestershire Police was one of the first police forces to trial the live Locate facial recognition technology from April 2014, notably using the tool to look for approximately 90,000 'known offenders' at the Download festival in June 2015.[149]

South Wales Police bid to the Home Office Police Transformation Fund in 2016, receiving £1,950,000 over two years (with the force committing £600,000 of their own funds) from January 2017 to deploy automated facial recognition in the context(s) of counter-terrorism; major events; body worn video; mobile phone app(s); automated number plate recognition;

[139] Davies B, Innes M, and Dawson A, *An Evaluation of South Wales Police's Use of Automated Facial Recognition* (Universities' Police Science Institute and Crime & Security Research Institute, Cardiff University 2018), 9; on automation versus augmentation more generally, see Veale M and Brass I, 'Administration by Algorithm? Public Management Meets Public Sector Machine Learning' in Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (Oxford University Press 2019).

[140] It should be noted that the two latter applications raise significant legal challenges in relation to the necessity of facial recognition for detecting or preventing crime.

[141] Written answer from Baroness Williams of Trafford, HL Deb, 28 July 2016, cW (HL1211).

[142] Ibid (HL1213).

[143] Wiles P, Annual Report 2017:Commissioner for the Retention and Use of Biometric Material (Office of the Biometrics Commissioner 2018), 88.

[144] In at least the case of South Wales Police, the system deployed is called 'NeoFace Watch'. See Davies B, Innes M, and Dawson A, *An Evaluation of South Wales Police's Use of Automated Facial Recognition* (Universities' Police Science Institute and Crime & Security Research Institute, Cardiff University 2018), 11.

[145] South Wales Police did however report lobbying NEC to include a gender filtering function on a new version of the the algorithm, which they then did. See ibid, 31.

[146] Nilsson P, 'How UK Police Are Using Facial Recognition Software', *Financial Times* (12 October 2018).

[147] The limit to this system is unclear, as Leicestershire Police reported applying this system with 90,000 records in the database to be live-scanned against. However, it was noted in the South Wales Police case that certain uses of this system left it slow and laggy. See Davies B, Innes M, and Dawson A, *An Evaluation of South Wales Police's Use of Automated Facial Recognition* (Universities' Police Science Institute and Crime & Security Research Institute, Cardiff University 2018).

[148] Ibid, 11.

[149] Purshouse J and Campbell L, 'Privacy, Crime Control and Police Use of Automated Facial Recognition Technology' (2019) 3 Criminal Law Review 188, 190.

and child sexual exploitation.[150] Funding was conditional on an evaluation of the technology being undertaken, which was carried out and published by Cardiff University.[151]

The force deployed the Locate system live at a number of sporting and music events in 2017/18. The threshold for inclusion as an individual of interest was unclear in several of these events to the evaluation team, although was noted to include strategies targeting pickpockets and mobile phone theft.[152] An initial deployment at the UEFA Champions League saw 2,632 matches, of which only 3% were 'true positives' and only one arrest resulted; later deployments saw the threshold for matching significantly increased, with a small music event, Elvisfest, matching 18 individuals with a 61% true positive rate, and a large boxing match matching 60 individuals with a 9% true positive rate.[153] A newer algorithm deployed later saw true positive rates ranging from 14 to 46%.[154] Newspapers have widely reported on these high false positive rates.[155] Where individuals were approached mistakenly, Cardiff University reported that "for the most part, interactions… were amicable […] operators / officers fully explained the exercise being carried out, and the individuals were invited into the vans to see the software for themselves and to see their own CCTV image alongside the 'match'".'[156]

The force deployed the Identify system on a laptop in the headquarters, with one member of staff a day responsible for facial recognition functions. Initially, many of the images sent from the field were of poor quality, such as mobile snapshots of CCTV, and significant organisational effort was required to train officers to ensure they were only sending high quality images.[157] Concerns with this system were that certain individuals were being repeatedly being matched to many photos, and officers reported that a commonality was that these individuals either had old photographs or had facial disfigurements.

The **London Metropolitan Police** undertook 10 deployments of live facial recognition technology between August 2016 at Notting Hill Carnival and February 2019 in Romford Town Centre.[158] Similarly to South Wales Police, a subset of individuals from the Metropolitan Police Service's databases of photographs were extracted, primarily drawn from photos taken while individuals were in custody but also, controversially, from other sources.[159] An evaluation undertaken by the University of Essex is forthcoming now the trial is complete.[160] However, as with the UEFA Champions League case undertaken by South Wales Police, the London Metropolitan Police has come under heavy fire for the revealed number of false positives, particularly in their Notting Hill Carnival deployment in 2016.[161]

[150] Davies B, Innes M, and Dawson A, *An Evaluation of South Wales Police's Use of Automated Facial Recognition* (Universities' Police Science Institute and Crime & Security Research Institute, Cardiff University 2018), 9, 12; Big Brother Watch report slightly different figures, reporting that South Wales 'was awarded a total of £2.6m by the Government to carry out automated facial recognition - £1.2m in 2016/2017 and £0.8m for 2017/18 by the Home Office, as well as £0.6m from Home Office Biometrics. South Wales has additionally contributed £100,000.' See Big Brother Watch, *Face Off: The Lawless Growth of Facial Recognition in UK Policing* (Big Brother Watch 2018).

[151] Davies and others, ibid.

[152] Ibid, 14.

[153] Ibid, 21.

[154] Ibid, 25.

[155] See e.g. Burgess M, 'Facial Recognition Tech Used by UK Police is Making a Ton of Mistakes' [04-05-2018] Wired.

[156] Ibid, 39.

[157] Ibid, 30.

[158] London Metropolitan Police, 'Facial Recognition to Take Place in Romford' (*London Met*, 13 February 2019) <http://news.met.police.uk/news/facial-recogntion-to-take-place-in-romford-358589> accessed 28 April 2019; London Policing Ethics Panel, *Interim Report on Live Facial Recognition* (LPEP 2018), 7.

[159] London Policing Ethics Panel, ibid, 14.

[160] Biometrics and Forensics Ethics Group Facial Recognition Working Group, *Ethical Issues Arising from the Police Use of Live Facial Recognition Technology* (Biometrics and Forensics Ethics Group 2019).

[161] Big Brother Watch, Face Off: The Lawless Growth of Facial Recognition in UK Policing (Big Brother Watch 2018);

The video recordings from the static cameras placed for facial recognition deployment for the London Metropolitan Police are retained for 30 days and then destroyed, with no images extracted.[162] Recognised individuals are retained until one month following the end of the trial, with only the evaluation team provided access.[163]

7.2.1 Concerns and legislation around data and privacy

As noted above, there are currently approximately 12.5m biometric, searchable faces on the Police National Database. Police forces have powers to take certain photos with or without consent, and to disclose or retain them for purposes relating to the prevention or detection of crime, the investigation of an offence or the conduct of a prosecution or to the enforcement of a sentence.[164] Such photos are permissible when individuals have been detained at a police station,[165] or under a range of other circumstances outside of a police station such as when they have been arrested, given a penalty notice or a direction under laws regulating antisocial behaviour.[166] Notably, there are no powers under this section of the Police and Criminal Evidence Act 1984 (PACE) to photograph and retain or repurpose images of individuals in public spaces more generally, and the main legal framework for these remains the Data Protection Act 2018.

The area of custody photographs has become an area of contention. In 2009, an anti-arms trade campaigner was photographed leaving a corporate meeting, and that photo was retained by police for broad intelligence purposes. The Court of Appeal held that there had been an unjustified inference with the claimant's right to respect for private life (art 8 ECHR), and, given that the man had not committed any criminal offences in the context of that image, such an image should have been deleted rather than indefinitely retained.[167] Some scholars have argued, however, that by making the 'reasonable expectation of privacy' test a touchstone as the Court did, the ruling significantly limits the protections to privacy foreseeable in public spaces, and places it out of step with both privacy scholarship and European Court of Human Rights (ECtHR) case law.[168]

[162] London Policing Ethics Panel, *Interim Report on Live Facial Recognition* (LPEP 2018), 7.

[163] Ibid, 14.

[164] Police and Criminal Evidence Act 1984 (PACE), s 64A(4).

[165] PACE s 64A(1).

[166] PACE s 64A(1B).

[167] *R (Wood) v Commissioner of Police of the Metropolis* [2009] EWCA Civ 414. See generally Grace J and Oswald M, '"Being on Our Radar Does Not Necessarily Mean Being under Our Microscope": The Regulation and Retention of Police Intelligence' (2016) 22 European Journal of Current Legal Issues.

[168] Purshouse J and Campbell L, 'Privacy, Crime Control and Police Use of Automated Facial Recognition Technology' (2019) 3 Criminal Law Review 188.

Facial Recognition Technology (FRT) and *Wood*

The main legal test case for determining which human rights are engaged by the use of FRT and overt public surveillance is *Wood*[1]. The claimant had argued that the police had violated his rights under Article 8 (right to the family and private life), 10 (freedom of expression), 11 (freedom of assembly and association) and 14 (non-discrimination in how ECHR rights are applied).[1] The claimant was photographed by police at a protest outside the Annual General Meeting of a company connected to the arms trade. He argued that the retention of these photographs by the police violated the aforementioned rights.

The Court of Appeal in *Wood* did not find that Articles 10, 11, and 14 had been breached, although they did not go into much detail as to why that was the case.

They did, however, find that his rights under Article 8 had been breached. While this particular application of Article 8 concerns the overt police surveillance in public spaces, rather than FRT directly, it is a useful reference point in understanding how FRT might engage similar rights. The Court of Appeal held that the activities of the Police interfered with the claimant's article 8 rights, and that this interference was not 'necessary in a democratic society', that is to say, the surveillance measures were disproportionate: the claimant "had not been ejected from the meeting", and was not guilty of any misconduct.

The Court of Appeal set out a three-stage process for determining whether Article 8 was breached[1]:

- The police measure must attain a "certain level of seriousness";
- It must be determined whether the claimant enjoys a "reasonable expectation of privacy"; and
- The application of article 8 may be curtailed by the scope of the justifications available to the state.

Because the police activity involved "a good deal more than the snapping of a shutter", such as the storing and processing of personal information, and were targeted specifically at the claimant, article 8 was found to have been breached.

In 2012, the High Court ruled in favour of two individuals who challenged the Metropolitan Police Service for retaining custody photographs taken under PACE.[169] This ruling stated that the "existing policy concerning the retention of custody photographs… is unlawful", and the police were given a "reasonable further period" for revising this policy – a period which should "be measured in months, not years".[170]

Significant contention has surrounded the seeming non-implementation of the ruling in this act. The Biometrics Commissioner was established in 2012, but the mandate of the role did not contain photographs.[171] Nevertheless, the annual reports from the Office of the Biometrics Commissioner consistently challenge photographs and the regulatory inaction and governance vacuums that are perceived to surround their use.[172] The House of

[169] R (RMC and FJ) v Metropolitan Police Service [2012] EWHC 1681.

[170] R (RMC and FJ) v Metropolitan Police Service [2012] EWHC 1681, para 58.

[171] Protection of Freedoms Act 2012, s 20.

[172] Marshall D and Thomas T, *Privacy and Criminal Justice* (Palgrave Macmillan 2017), 129.

Commons Science and Technology Committee too has raised this issue in their report on biometric data and technologies, noting that they are:

> "…particularly concerned to hear that the police are uploading photographs taken in custody, including images of people not subsequently charged with, or convicted of, a crime, to the Police National Database and applying facial recognition software. Although the High Court ruled in 2012 that existing policy concerning the retention of custody photograph by the police was "unlawful", this gap in the legislation has persisted."[173]

In February 2017, the government gave non-convicted individuals the right to ask police forces to delete their images from custody image database. A year later, 67 applications for deletion had been made, with only 34 successful.[174] This suggests that the current method for storing and deleting custody images is ineffective, and the approach stands in contrast to the millions of photographs stored in the Police National Database.

Furthermore, the fact that photographs are not simply labelled with the status of the individuals within it is problematic from a legal standpoint under the Data Protection Act 2018. There is a specific requirement under this statute in section 38(3) that:

> "In processing personal data for any of the law enforcement purposes, a clear distinction must, where relevant and as far as possible, be made between personal data relating to different categories of data subject, such as–
>
> (a) persons suspected of having committed or being about to commit a criminal offence;
>
> (b) persons convicted of a criminal offence;
>
> (c) persons who are or may be victims of a criminal offence;
>
> (d) witnesses or other persons with information about offences."

It appears the photographs being used have not been distinguished in this way, which could be argued to be an infringement of this law.

Regarding the specific facial recognition trials above, the Home Office has stated they are regulated in general by three regimes: the Data Protection Act 2018, the Surveillance Camera Code of Practice,[175] and general human rights principles.[176] None of these regimes specifically regulate facial recognition, and insofar as the ruling in *R (RMC and FJ) v Metropolitan Police Service* has not been implemented, specific questions concerning the rule of law in this area persist.

Whether such trials are indeed permitted by data protection law is an area of current contestation. South Wales Police's commissioned evaluation report notes that "it is not clear how [automated facial recognition] image use and retention fits with the new General Data Protection Regulation".[177] Big Brother Watch report in the case of South Wales that "biometric photos captured of at least 2,451 innocent people who have wrongfully been

[173] House of Commons Science and Technology Select Committee, Current and future uses of biometric data and technologies, Sixth Report of Session 2014-15, House of Commons: London, 25 February 2015, p.3.

[174] Press Association, ''Custody Image' Deletion Request Figures Revealed', *Mail Online* (12 February 2018) <http://www.dailymail.co.uk/wires/pa/article-5379353/Custody-image-deletion-request-figures-revealed.html> accessed 28 April 2019 (figures from 37/43 police forces in England and Wales, based on Freedom of Information requests).

[175] Established under the Protection of Freedoms Act 2012, s 29.

[176] Purshouse J and Campbell L, 'Privacy, Crime Control and Police Use of Automated Facial Recognition Technology' (2019) 3 Criminal Law Review 188, 198.

[177] Davies B, Innes M, and Dawson A, *An Evaluation of South Wales Police's Use of Automated Facial Recognition* (Universities' Police Science Institute and Crime & Security Research Institute, Cardiff University 2018), 40.

'matched' by facial recognition software remain in the hands of the police, entirely without their knowledge".[178] A judicial review is being sought by campaign group Liberty and a Cardiff resident against South Wales Police's use of facial recognition, while a separate review is being sought by Baroness Jenny Jones.[179]

Significant concerns can be identified when comparing the provisions of the UK's transposition of the Law Enforcement Directive 2016 in Part 3 of the Data Protection Act 2018 with police practices around facial recognition.

Facial recognition for law enforcement purposes constitutes sensitive processing under the DPA 2018, as it is biometric data processed for the purpose of uniquely identifying a natural person. When not relying on consent of the individual, processing this data must be strictly necessary for a law enforcement purpose, meet a condition in Schedule 8 of the DPA 2018, and be accompanied by an appropriate policy document.[180] It is highly unclear whether facial recognition at scale can meet a test of strict necessity, particularly given its highly unproven nature as discussed above.

Sub-Recommendation 5.2 Facial Recognition Datasets – Datasets used in facial recognition must operate clearly under the rule of law, adhering to conditions of strict necessity, and with categories of individuals clearly split as required under Part 3 of the Data Protection Act 2018. These must also specify how the data set has been selected to avoid selective sampling of the population, which could lead to bias and discrimination.

Sub-Recommendation 5.1 Facial Recognition Model Use – Facial recognition systems must operate clearly under the rule of law, with their lawful basis explicitly and openly defined, and this assessment should be made publicly available.

7.2.2 Concerns and legislation around bias

Big Brother Watch report that the software used by the UK police forces discussed above "has not been tested for demographic accuracy biases".[181] Such biases have been salient in the computing community, as commercially available facial recognition software has been shown to have error rates which differ based on demographic. Such systems have been demonstrated to perform poorly on Black individuals, and in particular on Black women.[182] Legally, the lack of analysis of such potential differences in the efficacy of the system could be construed as a failure to carry out the public sector equality duty in the procurement of this system[183] and a failure to consider the potentially discriminatory effects of profiling systems under GDPR[184] as well as potentially falling foul of other regimes were a system to specifically fail in a given case.

Beyond biases intrinsic to commercially available image recognition systems, where datasets and watchlists used do not simply consist of individuals who the police wish to charge with a

[178] Big Brother Watch, Face Off: The Lawless Growth of Facial Recognition in UK Policing (Big Brother Watch 2018), 29.

[179] Nilsson P, 'How UK Police Are Using Facial Recognition Software', *Financial Times* (12 October 2018).

[180] Data Protection Act 2018, s 35(8)(b)

[181] Big Brother Watch, Face Off: The Lawless Growth of Facial Recognition in UK Policing (Big Brother Watch 2018), 16.

[182] Buolamwini J and Gebru T, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' in *Conference on Fairness, Accountability and Transparency (FAT* 2018)* (2018).

[183] Equality Act 2010, s 149.

[184] General Data Protection Regulation 2016, recital 71 ("the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate […] that prevent, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect").

crime, a range of other problematic biases can be identified. For example, in 2017, the London Metropolitan Police included on their watch list those with mental health problems who 'fixated' on individuals in the public eye: a task that was accused of function creep and victimising the already societally marginalised, and which was carried out without consultation with mental health charities or stakeholder groups.[185] This could be considered further 'sensitive processing' under the Data Protection Act 2018, as it is processing data concerning the health of an individual, and goes above and beyond the biometric 'sensitive processing' already undertaken as part of facial recognition systems in general, which would require an identified and justified ground from Schedule 8.

7.2.3 Concerns and legislation around accountability, transparency and oversight

Many public bodies have reported on the lack of governance of police facial recognition. The Biometrics and Forensics Ethics Group, an advisory non-departmental public body sponsored by the Home Office, reported on police use of facial recognition systems and potential ethical frameworks, noting in particular the lack of independent oversight and governance.[186] The House of Commons Science and Technology Select Committee raised similar concerns.[187] The Information Commissioner launched an inquiry in December 2018 into police use of facial recognition technology,[188] having previously written a public-facing blog noting "how facial recognition technology is used in public spaces can be particularly intrusive" and that she was "deeply concerned about the absence of national level coordination in assessing the privacy risks and a comprehensive governance framework".[189] The Biometrics Commissioner has been consistently critical of the state of facial recognition, despite it falling outside of the role's statutory remit.[190] It is notable that the Home Office does not routinely make information about facial recognition using the Police National Database public, with information only appearing on request in relation to parliamentary questions or in an *ad hoc* manner to offices such as that of the Biometrics Commissioner.[191] The Commission recommends a strengthened Biometrics Commissioner, both in terms of statutory responsibilities and resourcing.

Sub-Recommendation 5.3 Biometrics Commissioner – The scrutiny powers, resources, and consultation role of the Biometrics Commissioner should be strengthened, and the scope of the Commissioner broadened and regularly reviewed.

It is further concerning that, as noted above, police forces appear to be unable to negotiate or scrutinise the technology they are deploying. Unlike some of the hotspot policing efforts discussed in section 7.1, facial recognition technologies have not been developed in-house or in academic consortia where the code and development is in public hands, but effectively

[185] Big Brother Watch, Face Off: The Lawless Growth of Facial Recognition in UK Policing (Big Brother Watch 2018), 27–28.
[186] Biometrics and Forensics Ethics Group Facial Recognition Working Group, *Ethical Issues Arising from the Police Use of Live Facial Recognition Technology* (Biometrics and Forensics Ethics Group 2019), 3.
[187] House of Commons Science and Technology Committee, Current and future uses of biometric data and technologies, Sixth Report of Session 2014-15, House of Commons: London, 25 February 2015, p.3.
[188] Hill R, 'ICO to Probe Facial Recog amid Concerns UK Cops Can't Shake Their Love for Unregulated Creepy Tech', *The Register* (3 December 2018).
[189] Denham E, 'Facial Recognition Technology and Law Enforcement', *The Information Commissioner's Office Blog* (14 August 2018) <https://ico.org.uk/about-the-ico/news-and-events/blog-facial-recognition-technology-and-law-enforcement/> accessed 28 April 2019.
[190] Marshall D and Thomas T, *Privacy and Criminal Justice* (Palgrave Macmillan 2017), 129.
[191] Wiles P, Annual Report 2017:Commissioner for the Retention and Use of Biometric Material (Office of the Biometrics Commissioner 2018), 88.

outsourced to the private sector. As with any area, there is a diversity within those developing solutions – some are robust and adhere to high standards (albeit self-set, and dependent on the use these technologies are put to), and those who do not – either because they have no incentive to be concerned, or they lack the capacity to make such a robust system. This is exacerbated by the desire to deliver products quickly, fast, and continue their development *in situ*, linked closely to ideas of quickly and cheaply developing a minimal viable product, and improving it through agile development.[192]

Determining which human rights are engaged, and how, is a complex exercise, but one which we argue is important. One approach would be for the suppliers engaged through public procurement processes to be required to conduct Human Rights Impact Assessments (HRIAs). An HRIA identifies, analyses and evaluates human rights considerations, in order to mitigate adverse impacts, and can be conducted at each stage of the design, development and deployment process. This has the dual effect of providing businesses a structured and guided governance framework to help support stronger decision making, whilst also empowering right holders to be able to better hold to account breaches. An *ex ante* framework, which seeks to avoid harmful consequences downstream, is far more likely to succeed in creating an 'ethics by design' environment and in turn more likely to be trusted.

The responsibility in the supply chain does not rest with the developers alone. The procuring party has a duty to ensure that it is buying in tools which meet the expected standards, equally as the procuring party it also wields significant power and influence over the supply chain, and the standards expected. The public sector has several avenues for ensuring private sector counterparts carry out HRIA, where proportionate to the end context. Legal strategies such as procurement frameworks, contractual clauses and certification models should be part of a range of measures that can contribute to the effective implementation of HRIA mitigation measures. As part of the procurement process, public sector buyers should consider the analytical and developmental capacity that any supplier has in relation to issues of human rights when engaging them – the robustness of such issues should be as relevant as the technical competence of the supplier. Allied to this, there ought to be full transparency between the supplier and purchaser to ensure access to all relevant information, assumptions and processes.

This effort should not just be carried out in criminal justice, as it opens up questions of human rights and values in a design context more broadly, considering the important role of upstream developers and providers on downstream services and policies. The Commission recommends a specific review be commissioned by government to consider policy responses at the intersection of human rights and technology.

Sub-Recommendation 4.2 Human Rights by Design – The Government should commission a review into policy options for mandating human rights considerations in technological design within different consequential sectors, including in the criminal justice system. This review should consider how and where human rights impact assessments should be required in public procurement processes.

[192] See generally Gürses S and Hoboken J van, 'Privacy after the Agile Turn' in E Selinger, J Polonetsky, and O Tene (eds), *The Cambridge Handbook of Consumer Privacy* (1st edn, Cambridge University Press 2018).

7.3 Individual Risk Assessment

Machine learning systems in policing can not only be applied to attempt to predict risk of certain events happening in particular places (section 7.1): they can also be used to predict the risk of particular individual exhibiting behaviours or characteristics in the future. A report from RAND on predictive policing tools highlight that both suspected offenders and potential victims can be subject to this form of risk scoring.[193] In this report, we add a third category of individuals to this list which is not discussed in the RAND report: police officers themselves, whose behaviour can be subject to prediction in an effort to manage forces and resource.

In England and Wales, individual risk scoring has been trialled and deployed in a variety of contexts, which will be examined in turn below.

7.3.1 Scoring at arrest

In the United States, much of the focus on algorithms in the public sector has been on their role in sentencing and in parole decisions.[194] In the United Kingdom, a parallel debate in the media can be seen earlier on in the justice pipeline: at the point where a decision as to whether and how to charge an individual who has been arrested is made. The highest-profile system of this type is the Harm Assessment Risk Tool (HART), which was developed in-house by Durham Constabulary in collaboration with the University of Cambridge in 2015/16 and deployed across the force at the point of custody decision.

HART's creation can be traced in part to the Turning Point programme undertaken by West Midlands Police and the University of Cambridge, an operation which sought to take certain vulnerable groups out of the traditional justice system and offer them alternatives to being charged for minor to moderate crimes – a practice known as 'out of court disposal'.[195] Durham Constabulary told the Commission that they were influenced by research that indicated women were treated more harshly "because decision-makers... subliminally considered that women also offended against their femininity".[196] Perceived success in initial trials focusing just on female offenders led to the creation of broader initiative aimed at a range of groups: Checkpoint. Checkpoint seeks "to tackle the root causes of offending" by "offering an alternative to prosecution for a very specific sub-set of criminal offenders".[197] It builds on the observation that prosecution for certain types of crime might itself fuel reoffending, and so for minor crimes such as the possession of drugs, a structured set of interventions based on a pathway model, in collaboration with organisations in sectors such as mental health or alcohol and drug dependency, is offered. Because the police have a six-month window in which they can charge following a crime, and because they wish to keep the threat of charging as a motivator for successful completion of the Checkpoint programme, these programmes are four months in duration. Fewer than 5% of individuals admitted to the Checkpoint programme fail it.[198]

While programmes like Checkpoint or Turning Point do not necessarily require algorithmic systems and interventions, the perceived need for the HART tool arose because of an

[193] Perry W, McInnis B, Price C, Smith S, and Hollywood J, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (RAND Corporation 2013), xiv.
[194] This debate was largely spurred by a 2016 investigation: Angwin J, Larson J, Mattu S, and Kirchner L, 'Machine Bias', *ProPublica* (23 May 2016).
[195] Coutts P, Turning Point: The Police's Production and Use of Evidence to Reduce Reoffending (Alliance for Useful Evidence, January 2018).
[196] Oral evidence to the Commission by Chief Constable of Durham Constabulary, Michael Barton (25 July 2018).
[197] Oswald M and others, 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and "Experimental" Proportionality' (2018) 27 Information & Communications Technology Law 223.
[198] Oral evidence to the Commission by Chief Constable of Durham Constabulary, Michael Barton (25 July 2018).

observation that custody officers found it challenging to identify individuals who have a moderate risk of recidivism. Ideally Checkpoint should not be offered to individuals with a low risk of recidivism, because as the aim is to reduce reoffending, it would be inefficient to use resources on cases where recidivism risk is low. It equally, and likely more importantly, should not be offered to high-risk individuals (those likely to commit serious offences such as murder, aggravated violent offences, robbery, sexual crimes, and firearm offences within the next two years) as not only would they be unlikely to be receptive to a Checkpoint-style intervention, but there would be a clear public safety risk from refraining from prosecuting.

The aim of HART within the Checkpoint programme is consequently to identify a middle stratum of risk where individuals do not need to be charged, and to reduce the number of people entering the justice system, and by doing so, hopefully reducing the number of people re-entering it. The idea that better predictive risk assessment can reduce the number of individuals detained in prisons is a common refrain heard from proponents of these system.[199] A review of studies recidivism prediction tools does indicate that the use of them improves clinical judgement from the perspective of predictive validity.[200] Custody officers, thought by Durham Constabulary to be unwilling or to find it challenging to accurately distinguish moderate risk individuals from low or high ones, would be supported by the HART tool. Chief Constable Michael Barton told the Commission that custody officers tended to be risk-averse, and err on the side of declaring an offender higher risk than data might justify.[201] Furthermore, at times where no skilled custody officer was on shift, the HART tool might play a larger role in supplementing decisions.

The HART tool is based on a random forest algorithm, which is a form of supervised machine learning.[202] Random forests are ensembles of decision trees grown on perturbed or sampled versions of the same dataset. Each of these trees – there may be thousands – might be using a different flowchart-like logic for prediction, and might capture a slightly different element or aspect of the phenomenon being modelled. Alone, any one tree could be too simplistic to understanding the complexities of the phenomenon, and may not include all the variables of interest as branches, but each of them is queried simultaneously, and together they ‘vote’ on the outcome. The outcome – such as low, medium or high risk – that the most trees vote for is the result of the random forest model. The random forest underpinning HART was initially trained on 104,000 custody events that occurred between 2008 and 2012, each of these events represented by 34 different predictors most of which concern offenders’ criminal behaviour histories.[203]

The HART tool received criticism for its use of i) postcode data and ii) personal data purchased from data broker Experian (data from their Mosaic programme). The use of location data in a policing context in general has been likened to historical redlining in the US, where due to housing segregation, home location can closely mirror ethnicity or other sociodemographics.[204] The use of data purchased about individuals in police databases also seems an overreach into individuals’ privacy and private life. Additionally, data brokers such as Experian have been accused of acting illegally in their amassing of these datasets, and

[199] See Eckhouse L, Lum K, Conti-Cook C, and Ciccolini J, ‘Layers of Bias: A Unified Approach for Understanding Problems With Risk Assessment’ (2019) 46 Criminal Justice and Behavior 185, 202–3.
[200] Desmarais SL, Johnson KL, and Singh JP, ‘Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings’ in *Handbook of Recidivism Risk/Needs Assessment Tools* (John Wiley & Sons, Ltd 2017).
[201] Oral evidence to the Commission by Chief Constable of Durham Constabulary, Michael Barton (25 July 2018).
[202] See generally Breiman L, ‘Random Forests’ (2001) 45 Machine Learning 5.
[203] See Urwin S, ‘Algorithmic Forecasting of Offender Dangerousness for Police Custody Officers: An Assessment of Accuracy for the Durham Constabulary Model’ (MSc, University of Cambridge 2016) Appendix B for a list of the relevant variables used.
[204] See Barocas S and Selbst AD, ‘Big Data’s Disparate Impact’ (2016) 104 California Law Review 671, 712.

are subject to several high profile complaints from organisations such as Privacy International.[205] For police forces to reuse such datasets when there is doubt about their lawful basis for existing seems problematic.

7.3.2 Scoring suspects

The **London Metropolitan Police** were provided with software from accounting firm Accenture at no charge while Accenture sought to increase its justice portfolio. This technology was aimed at creating a dashboard of individuals whom the force was concerned about in relation to knife crime. This dashboard featured a machine learning-created risk score alongside other features, such as a view on the individual's social media feed where applicable.

The **London Metropolitan Police** also created a 'Gangs Matrix', a forecasting tool designed to identify individuals with a propensity for either being a gang member or engaging in gang-related activities. London Metropolitan Police claim that this matrix is populated manually without the use of advanced algorithms or third-party software. This hybrid between a decision-support system and a database has however raised concerns in civil society: Amnesty International note that 87% of the people in the Gangs Matrix were from black and ethnic minority communities, while 78% were black. 75% were victims of violence themselves, and 35% had never committed a serious offence.[206]

Avon and Somerset Police have created statistical models concerning suspects' future behaviour, such as risk of perpetrating serious a domestic violence offence, a sexual violence offence, or a burglary. Around 250,000 potential (re-)offenders are given a score in this system.[207] Such models are built in-house on procured visual language platform for machine learning, IBM SPSS Modeller, procured in December 2013,[208] and the results of this models delivered to staff through a dashboarding system called QlikSense. Over 30 apps are deployed on the QlikSense visualisation platform by Avon and Somerset, and approximately 4,000 licences to use the software have been issued. The system was first piloted in 2016 in the context of £80m cuts to the force.[209]

The Commission also heard from Mike Edwards, Senior Lecturer in the International School of Policing and Security at the University of South Wales, about **South Wales Police** using algorithms to analyse the social media accounts of offenders, or potential offenders, although this was still in the exploratory phase.

7.3.3 Scoring victims

Some scoring systems take the unit of analysis as the potential victim of a crime. For example, **Avon and Somerset Police** have deployed predictive systems with regard to tackling the challenge of missing children. Individual records, such as report and phone calls, are used to create this system, but one of its tasks is predicting when children (in general) are likely to go missing, rather than specific children in the area. This is designed to ensure that staffing and rota levels are appropriate for these difficult-to-anticipate and time-

[205] Privacy International, Submission to the Information Commissioner–request for an assessment notice of data brokers Experian & Equifax (PI 2018).

[206] Amnesty International UK, Trapped in the Matrix: Secrecy, stigma, and bias in the Met's Gangs Database (2018).

[207] Dencik L, Hintz A, Redden J, and Warne H, 'Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services' (Data Justice Lab, Cardiff University 2018) 75.

[208] Aecus European Innovation Awards, *"Risk-based" predictive modelling to assist in the tackling of Burglary crime within Force Area* (2015) https://www.aecus.com/wp-content/uploads/2015/05/IBM-Avon-Somerset-Police_Risk-based-predictive-modelling-to-assist-in-the-tackling-of-Burglary-crime.pdf; Couchman H, 'Policing by Machine' (Liberty 2018) 46.

[209] Dencik L, Hintz A, Redden J, and Warne H, 'Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services' (Data Justice Lab, Cardiff University 2018) 74.

consuming activities. They have also built further models concerning potential victims, such as models predicting the propensity of being a victim of stalking and harassment. How – if at all – these models are deployed on the ground is unclear.

7.3.4 Scoring staff

Avon and Somerset Police additionally do some work where the police officers themselves are the subject of algorithmic analysis. Data quality issues are important in machine learning modelling, and police officers do not always record model-quality data. This is understandable, as in general, data in public services is secondary, collected for a very different task, and the manner in which it is collected is not always suitable for re-use.[210] Avon and Somerset use algorithmic systems to scan their datasets for potential errors, doing so also at an officer level to understand the sources of input errors and crime misclassifications which might affect their algorithmic systems further downstream.[211]

While not in the United Kingdom, another relevant use of officer-level predictive analysis was undertaken by a range of university researchers with police officers in North Carolina in the United States, who attempted to predict police officers at risk of committing misconduct in an effort from preventing it leading to serious consequences for themselves and others.[212] This can be seen in the broader context of the growth of technologies for employee surveillance, and a reminder that fundamental rights are at stake in relation to the staff within the justice system as well as those engaging with it in other ways.[213] Relatedly, work between University College London and the **London Metropolitan Police** has also focused on tracking the GPS units worn by police forces to shine light on the behaviour,[214] and the College of Policing have produced information on the use of body-worn cameras.[215]

7.3.5 Scoring in the prison system

The Offender Assessment System (OASys) is a national risk/need assessment tool with algorithmic components used across probation areas and prison establishments in England and Wales. Originally trialled as a paper-based system, OASys rapidly became a digital tool, and has gathered a range of digital components.[216] OASys was designed by the Ministry of Justice, and is today managed in close connection to the Ministry of Justice Data Science Hub. Its stated aims are to:

- assess how likely an offender is to reoffend;
- identify and classify offending-related needs;
- assess risk of serious harm, risks to the individual and other risks;
- assist with management of risk of serious harm;

[210] Veale M, Van Kleek M, and Binns R, 'Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making' in *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2018* (ACM Press 2018) doi:10/ct4s.

[211] Dencik L, Hintz A, Redden J, and Warne H, 'Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services' (Data Justice Lab, Cardiff University 2018).

[212] Carton S, Helsby J, Joseph K, Mahmud A, Park Y, Walsh J, Cody C, Patterson CE, Haynes L, and Ghani R, 'Identifying Police Officers at Risk of Adverse Events' in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '16, New York, NY, USA, ACM 2016).

[213] Edwards L, Martin L, and Henderson T, 'Employee Surveillance: The Road to Surveillance is Paved with Good Intentions' (SSRN Scholarly Paper, Social Science Research Network 18 August 2018).

[214] Kowalska K, Shawe-Taylor J, and Longley P, 'Data-Driven Modelling of Police Route Choice' in *Proceedings of the 23rd Conference on GIS Research UK 15th - 17th April, 2015 University of Liverpool, UK* (2015).

[215] College of Policing, *Body-Worn Video* (2014).

[216] See generally Mair G, Burke L, and Taylor S, '"The Worst Tax Form You've Ever Seen"? Probation Officers' Views about OASys' (2006) 53 Probation J. 7.

- link the assessment to the sentence plan;
- indicate the need for further specialist assessments; and
- measure change during the offender's sentence.[217]

OASys assessments are carried out at several points in the justice system: for a pre-sentence report; at the start of a sentence in prison or in the community; at regular review periods; at key decision points, such as when an offender is up for parole; and on termination of the sentence.

An OASys assessment provides 3 statistically validated indicators of reoffending:

- OASys General Predictor score (OGP) [non-sexual, non-violent offences]
- OASys Violence Predictor score (OVP)
- Offender Group Reconviction Score v.3 (OGRS3).

OGRS3 is a static prediction system computed only using static and non-changing components, such as demographic and criminal history. It is the easiest to calculate, because it requires the least data. OGP and OVP use the OGRS3 score as one component (although OVP gives higher weight to OGRS3's violence components) in addition to containing dynamic information. Both OGP and OVP use dynamic information on accommodation, employment, 'thinking and behaviour' and 'attitudes', while OGP uses lifestyle, associates and drug misuse, while OVP uses alcohol misuse and emotional wellbeing. This information is collated by probation officers or prison officers of different seniority appropriate to the case through document and file analysis alongside interviews to confirm its veracity. All OASys assessments are countersigned by a senior practitioner, with regular quality assurance activities carried out at a regional level.[218]

As all predictors share a common statistical underpinning in logistic regression, and a common component in OGRS, this section of the report will focus mainly on discussing the history, use and trajectory of OGRS.

Different forms of algorithms have long been in use in the prison system in the United Kingdom. An earlier version of OGRS has been in use since 1996 in England and Wales. It has not fundamentally changed: back then, too, it was a static, actuarial risk assessment instrument based on high level information about an individual – their age, gender and official criminal history. It predicts the percentage probability of criminal reoffending for individuals of adults discharged from custody or given non-custodial sentences within two years.[219] The initial development of OGRS relied heavily on the computerisation of the Home Office Offenders Index research database in the 1990s, which opened up a range of new analytic opportunities in relation to recidivism which had previously required laborious analysis of microfiche data.[220] OGRS1 set a template its successor scales would follow: the criminal records of tens of thousands of offenders were retrieved, and using logistic regression – a coarse form of machine learning – run against their determined recidivism status within two

[217] Debidin M (ed), A Compendium of Research and Analysis on the Offender Assessment System (OASys) 2006-2009 (Ministry of Justice 2009) 1-2.

[218] Moore R, *The Offender Assessment System (OASys): Development, validation and use in practice* <https://service.mvnet.de/_php/download.php?datei_id=38021>.

[219] Howard P, 'Offender Group Reconviction Scale' in *Handbook of Recidivism Risk/Needs Assessment Tools* (John Wiley & Sons, Ltd 2017).

[220] See generally Kershaw C, 'Interpreting Reconviction Rates' in *Papers from the British Criminology Conference, Queens University, Belfast, 15-19 July 1997* (The British Criminology Conferences: Selected Proceedings, British Society of Criminology 1999) vol 2.

years of discharge from custody or community sentence.[221] The logistic regression parameter estimates were rounded due to the limited technology available to probation staff on the front lines in 1992. For example, in an early version, being female was -3 points, male 0 points, and seven offence groups were scored ranging from -12 for a sexual offence to +6 for a drugs-related offence.[222]

By the time OGRS version 3 was developed, the Home Office Offenders Index had been replaced with a research version of the Police National Computer, which allowed more types and details around offences to be incorporated into the model. While developing this system, an earlier scoring system specifically for violent sexual offences, OGRS2SV, was shown to not be predictively valid and was withdrawn without replacement. The model was rationalised, removing three of the nine predictors required to ease the burden on assessors. It was only with OGRS3 that the model was used in prisons as well as in relation to probation, used as decision-support for sentence planning.

A core feature of all OGRS models (although the exact implementation changed over time) is the Copas rate, named after Professor John Copas of the University of Warwick who created OGRS1. The Copas rate is a parameter calculated on the basis of the length of years of the offender's known criminal career and their total number of convictions/sanctions. The rate is the highest when the offender has committed many offences in a short timespan. In OGRS4, the current version, the Copas rate is only used for offenders with three or more sanctions, as not to be biased against first-time offenders with short criminal histories.

Some efforts to tackle bias in these systems in the late 00's had already been occurring, largely before the academic work on bias detection.[223] Separate models were trained for women and men after it was realised that predictive validity was not as strong for female offenders as for male offenders: the result was that age was modelled separately for both.[224]

The current version of OGRS, OGRS4, was trained on 1,809,000 offenders released from custody or disposed of otherwise between Apr 2005 and Mar 2008 who had not reoffended before the end of March 2008, and recalibrated based on a further set of 174,000 offenders in 2010. OGRS4 input data consists of the following data from the previous seven years of records:

- Gender
- Age (at first conviction, sentence and release/order)
- Number of previous convictions (all offences)
- Number of previous sanctions for all offences, including convictions, cautions, reprimands and final warnings
- Type of offence (out of 20 categories)
- Current or previous breach
- Current or previous burglary

[221] Copas J and Marshall P, 'The Offender Group Reconviction Scale: A Statistical Reconviction Score for Use by Probation Officers' (2002) 47 Journal of the Royal Statistical Society: Series C (Applied Statistics) 159.

[222] Howard P, 'Offender Group Reconviction Scale' in *Handbook of Recidivism Risk/Needs Assessment Tools* (John Wiley & Sons, Ltd 2017) 231-2.

[223] Some of the earliest work on methods for machine learning and bias is Kamiran F and Calders T, 'Classifying without Discriminating' in (IEEE 2009) 2009 2nd International Conference on Computer, Control and Communication.

[224] Howard P, 'Offender Group Reconviction Scale' in *Handbook of Recidivism Risk/Needs Assessment Tools* (John Wiley & Sons, Ltd 2017) 232.

- Number of previous youth custodial sentences
- Number of sanctions for violent offences.

These primary risk factors are in some cases combined through pre-set formulae to create manual secondary risk factors, before being fed into a logistic regression.[225] In machine learning, this important step is called feature engineering.[226] OGRS4 uses interaction terms between several of these variables, notably between age and gender, and between gender and sanctions for violent offence. OGRS4 also contains an ability to calculate 'offence-free time', which places value on how many whole months from the time of score generation an offender (who may have been disposed outside of custody, such as through a community order) has not offended.[227] This allows for longer term planning and to relax the assumption that recidivism risk is constant, and that individuals not offending does not have a rehabilitative effect.

The Ministry of Justice have notably published all the model weightings of this recidivism scoring system publicly, although the Commission is not aware of scrutiny or analysis of OGRS undertaken by any civil society bodies or NGOs in this area.[228] In the same document, they have also published predictive validity on a variety of demographic subgroups, as well as offenders broken down by offence type. They additionally have published both in-house and peer-reviewed work analysing and explaining disparities in predictive performance by age, gender and ethnicity.[229] They do note that it is resource intensive to publish model weights constantly, because the models require constant retraining and recalibration, but note for interested parties:

> "In reporting these refitted models, the model coefficients have not been reported. This acknowledges the intention for the predictors to be recalibrated frequently, ideally on an annual cycle, which will make the repeated publication of model coefficients overly resource-intensive. Instead, parties interested in using the predictors should contact NOMS (National.Research@noms.gsi.gov.uk) to gain information on the latest versions and discuss licensing matters as appropriate to the nature of their intended use."[230]

While the OGRS score is based on relatively basic machine learning, logistic regression, which does not include many variables nor by default have those variables interact with each other, the Ministry of Justice is considering whether further, more advanced machine learning methods, such as random forests, stochastic boosting, or ensemble methods, could be used, as well as whether these methods would allow the number of risk factors involved to be increased.[231] Indeed, reports from the Ministry of Justice indicate that they have been using neural networks and ensembles of trees in-house since as early as 2009 for better

[225] R Moore (ed), *A Compendium of Research and Analysis on the Offender Assessment System* (Ministry of Justice Analytical Series 2015) <https://perma.cc/W2FT-NFWZ> 158.

[226] Domingos P, 'A Few Useful Things to Know about Machine Learning' (2012) 55 Communications of the ACM 78.

[227] R Moore (ed), *A Compendium of Research and Analysis on the Offender Assessment System* (Ministry of Justice Analytical Series 2015) <https://perma.cc/W2FT-NFWZ> 153.

[228] They are available in R Moore (ed), *A Compendium of Research and Analysis on the Offender Assessment System* (Ministry of Justice Analytical Series 2015) <https://perma.cc/W2FT-NFWZ> 171-2, table 8.5.

[229] Howard P, 'The Effect of Sample Heterogeneity and Risk Categorization on Area Under the Curve Predictive Validity Metrics' (2017) 44 Criminal Justice and Behavior 103; Debidin M (ed), *A Compendium of Research and Analysis on the Offender Assessment System (OASys) 2006-2009* (Ministry of Justice 2009) 101.

[230] R Moore (ed), *A Compendium of Research and Analysis on the Offender Assessment System* (Ministry of Justice Analytical Series 2015) <https://perma.cc/W2FT-NFWZ> 294.

[231] Howard P, 'Offender Group Reconviction Scale' in *Handbook of Recidivism Risk/Needs Assessment Tools* (John Wiley & Sons, Ltd 2017) 239.

understanding the functioning and limitations of their simpler models.[232] It is also worth noting that a gender-binary view could in the near future find itself so insufficient as to not be useful.

7.3.5.1 Use of the OGRS score.

Like many state-provided services, the prison service takes a risk-based approach to its public tasks.[233] Individuals are managed at one of seven levels of 'service tier', with those at higher tiers managed by more senior staff, and/or met with more frequently. The OGRS score partly determines this tier: offenders with OGRS scores at or above 75% will be managed at or above the third-highest tier; those with scores of 50-74% at or above the fifth-highest tier. More granular determinations of tier are then organised by structured professional judgement, such as the individuals' risk to themselves and to others.[234]

The OGRS score also is provided as part of the pre-sentence report. Following the Criminal Justice Act 2003, a pre-sentence report is a report:

> "a) with a view to assisting the court in determining the most suitable method of dealing with an offender, is made or submitted by an appropriate officer
>
> b) and [which] contains information as to such matters, presented in such manner, as may be prescribed by rules made by the Secretary of State."[235]

Pre-sentence report date back to the 19th Century. Today, they are generally prepared in the weeks of adjournment between conviction and sentence.[236] As the name would suggest, they are designed to inform the sentencing process. They are also drawn upon later, in the parole process, if available.[237] They have a third, unofficial role, as "[p]ractice has shown that… [the pre-sentence report] can also form a critical part of the offender management process by providing insight to an individual's offending behaviour".[238] Research has a gap here, with academic research having gone so far as to note that "we currently know nothing about how [pre-sentence] reports are perceived or used in this context [of the supervising officer who inherits the case]".[239] In particular, the way that these users understand and are influenced by the OGRS score is unclear.

The OGRS score is not just part of pre-sentence report, but is used to rationalise the provision of report-writing resources in the justice system. Dr Gwen Robinson notes:

> "the developing [OGRS] technology of risk assessment was also seen as providing a defensible basis for the rationing of report-writing resources: a 2004 circular included a new 'same day' report format for use in cases where low risk was indicated. 'Low risk' at this time was defined as an Offence Group Reconviction Scale (OGRS) score of less than 31 per cent and an OASys risk of harm screening which did not indicate the need for a full risk of harm assessment."[240]

[232] Debidin M (ed), A Compendium of Research and Analysis on the Offender Assessment System (OASys) 2006-2009 (Ministry of Justice 2009) chapter 9.

[233] See generally Black J, 'The Emergence of Risk-Based Regulation and the New Public Risk Management in the United Kingdom' (2005) 3 Public Law 512.

[234] Howard P, 'Offender Group Reconviction Scale' in *Handbook of Recidivism Risk/Needs Assessment Tools* (John Wiley & Sons, Ltd 2017) 239.

[235] Criminal Justice Act 2003 s 158(1).

[236] Robinson G, 'Stand-down and Deliver: Pre-Sentence Reports, Quality and the New Culture of Speed' (2017) 64 Probation Journal 337, 338.

[237] The Parole Board Rules 2016.

[238] Ministry of Justice, *Determining Pre Sentence Reports* (PI 05/2011) (2011).

[239] Robinson G, 'Stand-down and Deliver: Pre-Sentence Reports, Quality and the New Culture of Speed' (2017) 64 Probation Journal 337, 348.

[240] Ibid, 339.

The Ministry of Justice has stated that pre-sentencing reports should “be in an appropriate format commensurate with the seriousness and complexity of the offence”.[241] The Criminal Justice Act 2003 removed the need for pre-sentence reports to be written,[242] while guidance maintained that the OGRS score should be communicated orally. A new short report format, designed to be turned around in five days, was prescribed for individuals with OGRS scores of under 41%.[243] In 2007, the ‘low risk’ threshold was again raised to an OGRS score of below 76%.[244] Importantly, only OGRS results over this threshold will go on to trigger the full OASys profiling in a pre-sentence context: the OASys OGP and OVP scores will not be calculated (at this stage) for an offender deemed low risk by these thresholds.

Frontline workers, who may be using OGRS scores in allocating resources or in preparing pre-sentence reports, are given formal training on OGRS and other risk predictors used in the prison service. They are taught what the scores are calibrated on, and are trained to accurately input information on two case studies in order to pass the course. The Ministry of Justice Data Science Hub maintains a spreadsheet which converts approximately 2,300 legally distinct, named and numbered criminal offences into the 20 categories used in OGRS4, and updates this annually as statutes change. Three phone lines are available to frontline workers who need assistance or have questions, focussing on ICT-related problems, offender management challenges, and data science respectively.[245]

7.3.6 Concerns and legislative framework

No legal provision explicitly requires the creation of actuarial or predictive risk scores on offenders. This is in contrast to some state laws in the US, such as SB-10 in California, which have required US counties to procure predictive recidivism systems.[246] However, the large body of sentencing law in force in England and Wales, which will not be reviewed in this report, does require the courts to consider different types of risk at many points in the justice system.[247] Pre-sentence reports, already discussed, are one provision in statute designed to lay information on risk before courts.

There are several means by which the Government could process data about offenders under the Data Protection Act 2018. Firstly, Schedule 7 of the DPA18 includes within its list of competent authorities, which fall under the Law Enforcement Directive/Part 3 DPA18, “authorities with functions relating to offender management”. These include private actors contracted to run prisons. Insofar as data processing can be construed as for the “purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security”, processing can happen under Part 3 of the Act. If processing under this part, rights are restricted. As discussed in this report, the rights for information about algorithmic decisions are removed.[248]

When data is processed under this Part however, some obligations are heightened. One of these is logging, where logs must be kept for any act of:

- collection;

[241] Ministry of Justice, Service Specification for Assessment & Reports Pre-Sentence (HM Government 2010).

[242] Criminal Justice Act 2003 s 158(1A).

[243] Robinson G, ‘Stand-down and Deliver: Pre-Sentence Reports, Quality and the New Culture of Speed’ (2017) 64 Probation Journal 337, 339-340.

[244] Ibid, 340.

[245] Howard P, ‘Offender Group Reconviction Scale’ in *Handbook of Recidivism Risk/Needs Assessment Tools* (John Wiley & Sons, Ltd 2017) 236.

[246] See generally Partnership on AI, *Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System* (2019).

[247] See generally Law Commission, Sentencing law in England and Wales: Legislation currently in force (2015).

[248] See section 6.1.

- alteration;
- consultation;
- disclosure (including transfers);
- combination; or
- erasure.

This level of logging is not required under GDPR, but under Law Enforcement Directive this would likely mean that detailed metadata about the use of personal data in training, assessing or deploying algorithmic systems must be retained. These logs are not intended to be provided to data subjects, but must be provided to the Information Commissioner upon request.

Such logging requirements are welcome, but it is unclear how terms such as ‘alteration’ and ‘consultation’ apply to algorithmic systems. Consequently, the Commission recommends the Information Commissioner provide guidance on how these important tasks are carried out, to ensure they apply to algorithmic systems with the rigour required to ensure these systems’ legality.

Sub-Recommendation 2.2 ICO Guidance on Logging for Algorithmic Systems – The ICO should provide guidance on how the logging requirements in Part 3 of the Data Protection Act apply to the use of algorithmic systems falling under this Part.

Some processing might also be carried out if “processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller”.[249] It could also be that, in the realm of offender management, which attempts to safeguard the general public as well as the individual, “processing is necessary in order to protect the vital interests of the data subject or of another natural person”.[250] Such processing could include the risk score creation, which the GDPR would characterise as profiling.[251] However, as *lex specialis*, the Law Enforcement Directive/Part 3 DPA18 must apply if a competent authority is processing data for law enforcement purposes – a competent authority cannot opt out of it and opt into GDPR instead.

Furthermore, the Data Protection Act 2018 limits the use of data subject rights, such as the right to access information, object to processing, or have data erased, for GDPR purposes involving the prevention or detection of crime or the apprehension or prosecution of offenders,[252] and comparatively to the GDPR, for processing by competent authorities for law enforcement purposes.[253] The Data Protection Act 2018, like the Data Protection Act 1998, does also specifically restrict the transparency and access provisions of the General Data Protection Act for risk assessment systems, however the risk assessment systems in scope only concern those related to housing benefit, taxation, and the unlawful use of public money.[254] Notably – even though it is allowed to be restricted under the GDPR[255] – Article 22, on automated decision making, is not restricted under these provisions. It may be

[249] GDPR art 6(1)(d).
[250] GDPR art 6(1)(e).
[251] GDPR art 4(4).
[252] Data Protection Act 2018 sch 2 para 2.
[253] See section 6.1.
[254] Data Protection Act 2018 sch 2 para 3.
[255] GDPR art 23.

possible for an offender to challenge the automated use of OGRS scores on this basis, such as the use of them in tiering resourcing. This would be challenging, however, because within that setup, as described above, a frontline worker makes the final granular assessment of which tier to fall in, and it is unlikely that this could be construed, as the GDPR requires, as a decision 'based solely' on automated processing.[256]

In theory, the restriction of the rights in the previous section would limit an offender's ability to inquire about the nature of the score. While the Ministry of Justice, as noted above, does publish the weights of the score,[257] they are not under any clear legal obligation to do so. Given that Article 15, the right to access, is restricted where data is processed for the purpose of the apprehension or prosecution of offenders, any hope of getting a general description of the model under data protection law by appealing to 'meaningful information about the logic of processing' seems difficult, even if it were accepted the preconditions to using that right were otherwise met.[258] If the Law Enforcement Directive applies, this right disappears entirely. In particular cases, it may be possible to draw upon the common law duty to give reasons, but given the lack of a general duty to give reasons, this is likely to be a fragile approach and remedy.[259]

Overall, it appears that the OGRS scheme and the surrounding infrastructure is institutionally strong, but in statutory terms, weak. There appears to be a strong research and reflection base that has long been considering issues such a bias and on-the-ground deployment, but the longevity of these efforts is unclear. Rather than an 'if it ain't broke, don't fix it' approach, the practices developed to date must be carefully incorporated as mandatory requirements.

Sub-Recommendation 6.1 Formalise Governance of Risk Scoring – The Government should take stock of the practices surrounding the development of risk assessment tools used in sentencing and offender management, and enshrine at least the current best practices – such as regular analysis, reviewing and reporting – as statutory responsibilities.

7.4 Digital Forensics

Digital forensics are a set of intelligence and evidential tools which involve applying scientific methods to the recovery, analysis and interpretation of relevant digital materials and data in criminal investigations and court proceedings to assist in delivering justice.

Digital forensics encompass a range of aims, including to identify or match individuals to digital materials or traces; interpret ambiguous digital materials or traces; reconstructing

[256] See generally Veale M and Edwards L, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling' (2018) 34 Computer Law & Security Review 398.
[257] R Moore (ed), *A Compendium of Research and Analysis on the Offender Assessment System* (Ministry of Justice Analytical Series 2015) <https://perma.cc/W2FT-NFWZ> 171–2, table 8.5,
[258] GDPR art 15(1)(h). Whether this provision would provide information on the logic of the whole model is subject to academic debate, and is unlikely to be resolve until placed in front of a court. Claiming that the right is limited, see Wachter S, Mittelstadt B, and Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76; cf Selbst AD and Powles J, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233. For an overview, see Kaminski ME, 'The Right to Explanation, Explained' (2019) 34 Berkeley Technology Law Journal __.
[259] See generally Elliott M, 'Has the Common Law Duty to Give Reasons Come of Age Yet?' (University of Cambridge Faculty of Law Research Paper No. 7/2012, University of Cambridge 2012); in relation to public sector algorithmic systems, see Oswald M, 'Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power' (2018) 376 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20170359.

whole events from a range of evidence; and offering broad opinion on a set of circumstances.[260] Furthermore, standard forensic practices, such as DNA testing and audio-visual analysis, use algorithmic methods.

This Commission did not focus on digital forensics, and regrettably did not receive significant numbers of submissions relating to it. We acknowledge the recent House of Lords Science and Technology Committee report, which touches on many of the challenges in governing this sector.[261] Furthermore, much of forensic science crosses a very wide array of disciplines, and despite the importance of algorithmic elements, the computing focus of this Commission is a narrow lens to cast on a problem that intersects with many wider issues, such as the biological sciences.

Despite this, the House of Lords pointed to a number of challenges which resonate with issues discussed earlier and further below in this report.

7.4.1.1 Concerns and legislative framework

The United Kingdom Accreditation Service (UKAS) is the sole accreditation body for forensic services in England and Wales. Providers to police forces must certify under ISO 17020 and ISO 17025 standards. While standardisation is welcome, and goes beyond what is seen in some of the other cases discussed in this report, these appropriateness of these standards for digital forensics has divided opinion, with some practitioners noting that many more suitable ISO standards for digital evidence are available.[262] In general, despite assurances that legislation would be introduced to give statutory powers to the Forensic Science Regulator, the Government has not done so, leaving it without important powers such as powers to rescind accreditation, limit individuals' ability to provide expert testimony, investigate and take enforcement action against forensic science providers, issue improvement notices and fines, and inspect providers without notice.[263]

Many of the concerns about opacity of algorithmic systems in the justice sector more widely are echoed in the area of digital forensics. Sir Brian Leveson told the House of Lords Science and Technology Select Committee:

> "…a commercial provider managed to download or retrieve some of the [messages from a phone which had been wiped]. The defence wanted to know how they had done that and the scientist was not prepared to explain it, first, because it was commercially confidential and, secondly, if he explained how he had done it, the next time round they would find a way of avoiding that problem."[264]

Obvious challenges around ensuring the efficacy and fairness of the technologies used and enabling evidence to be challenged on its merits surround this approach. These challenges are only likely to become greater as more and more crimes have significant sources of evidence in the digital domain.

We can also raise concerns around digital forensic tools in the context of the right to a fair trial. One key element of fair trials guaranteed by Article 6 of the European Convention on Human Rights is the notion of 'equality of arms'. Equality of arms means that if a person

[260] Sommer P, Written Evidence FRS0009 to the House of Lords Science and Technology Inquiry into Forensic science and the criminal justice system (2018).

[261] House of Lords Science and Technology Committee, *Forensic science and the criminal justice system: a blueprint for change* (HL Paper 333 2019).

[262] Ibid paras 81–2.

[263] Ibid para 109.

[264] Ibid para 148.

wishes to contest an accusation, charge or claim, “she must be in a position to argue her case on the basis of equal access to relevant knowledge in comparison to the prosecutor or claimant”.[265] Insofar as individuals in a legal process are unable to understand and contest, even with the help of legal counsel, complex algorithmic systems used to process evidence alleged to relate to them, there is a significant threat to due process rights and this ‘equality’ or ‘parity’ of arms that arguably needs to be supported.

This is particularly worrying as privatised tools and systems are the norm in the forensics sector, particularly as police forces with specialist digital units are overwhelmed and faced with a significant backlog.[266]

This area is in need of further investigation, and the Law Society looks forwards to working with Government, relevant regulators and other stakeholders following the release of the House of Lords Science and Technology Committee report, as well as within this fast-developing context more broadly. At this point, the Commission is mostly concerned around the capacity of the current digital forensics system to rigorously understand and keep on top of challenging issues.

Sub-Recommendation 6.4 Digital Forensics In-House Capacity – The Government must ensure that the public sector maintains significant, effective capacity to rigorously understand digital forensic issues.

7.5 Modernisation of the Courts

The HM Court and Tribunal Service’s (HMCTS) modernisation programme is a £1.2bn programme running from 2016 until 2023 and is designed, in part, to bring much needed modern digital technologies into the infrastructure and processes of the courts and tribunal services. The programme is ambitious, the scale of change required is daunting, and as a consequence, timescales have been recently amended to extend the programmes timeline.

The reforms, once in place, aim to introduce specific digital innovations. Litigants will be able to start proceedings online in the civil and family courts. There will be a process for tracking appeals in tribunals. Many hearings may take place via video link, including some civil applications and remand hearings in the criminal courts. For some criminal cases, it will be possible to plead guilty online, and receive an automatic statutory online conviction for offences such as non-payment of television licences. Some systems, such as digitising the transmission of information within courts, are likely to cross all domains. These systems, as of current plans, are likely to only be simplistic rule-based algorithms.

A significant strand of the reform programme is the overhaul of the court and tribunal estate. Many courts and tribunal centres are not fit for purpose after years of neglect and would be too costly to upgrade to allow for new technologies. The court closure programme, which in part is designed to release capital for reinvestment in the modernisation programme, has led to the risk of near-term harm in access to justice terms – closing courts before the technology that is intended to replace them is in place, tested and proven.

The Law Society believes that recommendations within this report, particularly those concerning areas such as procurement, auditing and transparency, also have relevance for

[265] Hildebrandt M, Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology (Edward Elgar 2015) 101.

[266] House of Lords Science and Technology Committee, *Forensic science and the criminal justice system: a blueprint for change* (HL Paper 333 2019), para 147 (reporting that ‘the Metropolitan Police had “a seven-month backlog”’).

the modernisation of courts going forwards. Professor Richard Susskind told the Commission:

> "Online Courts in their first generation don't at all involve AI making judicial decisions. […] But we can anticipate a second stage where some decisions may indeed – and this will be many years hence, but now is the time to start thinking about the implications of these - we can imagine a situation where some decisions might be made by machines rather than human beings."

Creating the right framework now, with clear standards and expectations, will pave the way for a more reasoned and robust design of any such future systems, avoiding the pitfalls evidenced in other areas. As Professor Susskind asks, "whether a court is a place or a service" either way, it is vital that all users of the courts trust their ability to be fair, open and objective.

We will be continuing to closely follow the work in this area, particularly in respect to any role for modelling or sophisticated data analysis within this programme. The data which sits within the courts system, albeit broadly unstructured at the moment, has the potential to offer valuable insights and policy evidence as to the impact and effectiveness. It is not inconceivable that in time this is capable of being utilised more fully. Ensuring that the core principles which arise from this report sit as a foundation layer to the development of the HMCTS modernisation programme would be an important way of futureproofing the courts.

7.6 Mobile Device Extraction

Another use of algorithmic techniques in criminal justice has surrounded the use of them to seize electronics information from computers and personal devices. In recent years, controversy has emerged over the extraction of information from the mobile phones of suspects, convicts, and victims. These practices have been called 'mobile device extraction', 'digital stop and search', 'digital device triaging', 'digital forensic triage', among other names. These involve the use of proprietary zero-day attacks, algorithmic methods of compromising the security of a device by relying on bugs that developers of devices are unaware of or unable to fix, and are operated through both self-service kiosks within police stations and central 'hubs'.

Privacy International identify 26 forces in the UK using this technology and raise six core questions concerning the interaction of mobile device extraction technologies with individuals' rights and freedoms:[267]

> "(1) Whether victims, witnesses and suspects, including those released without charge or found innocent, are aware that personal information may have been taken from their phones without their knowledge.
>
> (2) If consent is given by the user to the police force to extract data from mobile phones, how informed is that consent;
>
> (3) What happens to the vast amount of data that is copied from the device;
>
> (4) Whether data is shared with other bodies;
>
> (5) If this data is deleted, and if so, after how long; and
>
> (6) How securely the data is stored."

[267] Privacy International, *Digital Stop and Search* (PI 2018) 28.

Privacy International have pointed to the lack of reports or reviews of legality and process around these tools across the country, as well as a lack of local guidance on how to use these technologies appropriately and legally.

Mobile devices hold a great deal of invasive data. Given their increased power, storage, and use as a hub for many goods and services, mobile devices reach into almost every area of private and social life. There is arguably no clear comparison to the invasiveness of the centralisation of information on these devices, or the information they are authenticated to access. Consequently, it is clear an individual's right to private life under the ECHR is engaged, and that the relevant tests must be considered carefully to ensure that no unlawful infringement occurs.

Algorithmic techniques are both used to access and to analyse this data. For example, Privacy International report that the company Cellbrite offers forces functionality to "[i]dentify and determine the strength of connections between people, places and events by viewing maps and timelines" within its suite of mobile phone extraction tools.[268] They are also used to go beyond user expectations around the security of their devices, as they access information despite assurances from firms that this data is securely encrypted, even from the hardware manufacturer or software developer.

Interfering with rights under Article 8, ECHR requires the interference to be "in accordance with the law", the test of which requires a basis in adequately accessible, reasonably foreseeable domestic law, which itself is compatible with the rule of law.[269] Analysing the case in Scotland, some recent work has cast doubt on whether the interference caused by these devices is justified.[270] The lawful basis claimed by different police forces for this is patchy and inconsistent.[271] Further issues with mobile extraction are likely to emerge around articles 6 (right to a fair trial), 7 (no punishment without law), 9 (freedom of thought, conscience and religion) and 10 (freedom of expression),[272] as well as potentially touching upon legally privileged information.

Furthermore, looking at the case law of the European Convention of Human Rights, it appears that the ECtHR is of the view that prior judicial authorisation in the form of a warrant is required for the search of electronic devices to be compatible with the rule of law.[273]

The Information Commissioner's Office has also expressed concerns around the data protection compliance of these systems and practices, particularly in relevance to the wholesale and untargeted nature of the information retrieval. They stated to Scottish Parliament (but in relation to UK law):

> "If the police went through all of someone's text messages, that would potentially be an intrusion into other people's private conversations that were not relevant to the case; it would not simply be a case of focusing on the conversations between the particular persons who were already of interest. If that kind of interrogation leads to other people of interest, that evidence

[268] Ibid, 15.

[269] *Silver and Others v. United Kingdom*, App no. 5947/72; 6205/73; 7052/75; 7061/75; 7107/75; 7113/75; 7136/75 (ECtHR, 25 March 1983) paras 86-88.

[270] Rice M, 'Seizing the future: Seeking clarity of law in the search and seizure of mobile devices in Scotland', In: Proceedings of the 2019 British and Irish British and Irish Law Education and Technology Association Conference (BILETA 2019).

[271] Privacy International, *Digital Stop and Search* (PI 2018) 21.

[272] The Scottish Parliament, Justice Sub-Committee on Policing, Report on Police Scotland's proposal to introduce the use of digital device triage systems (cyber kiosks) (April 2019) para 126.

[273] See *Iiiya Stefanov v. Bulgaria*, App no. 65755/01 (ECtHR, 22 May 2001); *Prezhdarovi v. Bulgaria,* App no. 8429/05 (ECtHR, 30 September 2014).

> would be of further relevance to the case, but extracting everything wholesale in that way puts the police at a risk of non-compliance."[274]

There is also a concern around the existence of consent as a lawful basis in Part 3 of the Data Protection Act. Consent has been used by police forces to justify data extraction from mobile devices of suspects and victims.[275] This is problematic, as the Law Enforcement Directive indicates:

> "Member States shall provide for processing to be lawful only if and to the extent that processing is necessary for the performance of a task carried out by a competent authority for the purposes set out in Article 1(1) and that it is based on Union or Member State law."[276]

The UK's transposition of this Directive adds a legal condition for processing for law enforcement purposes not present in the original: consent.[277] The reliance on consent for data processing for law enforcement purposes is concerning to the Commission, as it seems unlikely that such consent could be freely given or withdrawn in situations of power imbalances, even by victims.

This Commission will not weigh in in detail around the legality of these technologies; suffice it to say that there appears to be legal uncertainty around the use of these tools in relation to individuals' rights and freedoms. This uncertainty leads to the Recommendation to further consider these issues.

Sub-Recommendation 5.4 Mobile Device Extraction Assessment – An appropriate body – potentially Her Majesty's Inspectorate of Constabulary and Fire and Rescue Services (HMICFRS) – should be tasked with establishing a working group to consider issues around the legal, effective and legitimate use of technologies to search seized electronic devices.

The ICO is considering UK law enforcement's use of 'Digital Triage Devices' and a related report is expected later in the year.[278] However, given that such a report will be unlikely to consider all aspects of relevant law, such as those beyond the remit of the ICO, the Commission feels it is important for multiple efforts to examine this challenge from different angles.

[274] The Scottish Parliament, Justice Sub-Committee, *Official Report* (September 13 2018) col 12.
[275] BBC News, 'Rape victims among those to be asked to hand phones to police' (BBC 29 April 2019).
[276] Law Enforcement Directive, art 8(1).
[277] Data Protection Act 2019 s 35(2)(a).
[278] Johnson S, Digital Triage Devices (Cyber Kiosks) (Scottish Police Authority, 8 May 2019), para 4.2.

8 Cross-cutting Rights and Issues

In this concluding section, the Commission will reflect on a range of issues that arose from both the individual cases of algorithmic systems in the context of wider societal systems, and common themes from multiple algorithmic deployments in and around the criminal justice system.

8.1 Lawfulness

UK, English and Welsh law already has a range of provisions that touch upon the use of algorithmic systems directly and indirectly. Many of those have been discussed above. There have been concerns aired in this report, particularly around facial recognition and mobile device extraction, that some existing systems sit uncomfortably with existing domestic and international legal provisions and obligations, especially our human rights framework.

The Commission has highlighted the importance of placing algorithmic systems within the existing legal regimes. It is simply not the case that the law is fundamentally incompatible with these technologies, as can occasionally be heard. It may be the case that more or less permissive regimes are desirable, but that is broadly a matter for Parliament to determine.

In most cases, this will require actors in the criminal justice system to, at the very least, elaborate upon the lawful basis of all algorithmic systems likely to touch upon legal issues, rights and freedoms. The confusion that ensues when this is not done is illustrated in the sections above on Facial Recognition in Policing and Mobile Device Extraction.

Interferences to Article 8 of the ECHR, for example, must respect the rule of law. If a lawful basis is unclear, it is very hard to say that it is being respected.

Recommendation 5 Lawfulness – The lawful basis of all algorithmic systems in the criminal justice system must be clear and explicitly declared in advance.

It is not always easy to for technological practices to keep up with the law, nor for the law to keep up with technology. This is not largely because entire frameworks, such as human rights or data protection, are inherently incompatible, but because the policy instruments may cease to effectively govern a complex system, or approaches taken may no longer strike a fair balance or appropriate trade-off in a changing world. As a result, where new technologies are at stake, or provisions are drafted with these in mind, it may be worth considering the role of adaptive governance. Governing and rulemaking with planned adaption considered allows for some of the assumptions made during the policy-making process, such as the accuracy of recidivism prediction systems, to be questioned as new evidence arises. It is used with relative frequency in some areas of law, such as environmental regulations, but has not seen much uptake to date in regimes governing computing systems in society.[279] One common technique is the sunset clause, where a provision exits force after a certain date unless some actions are taken concerning its continuation or revision.

[279] See eg McCray LE, Oye KA, and Petersen AC, 'Planned Adaptation in Risk Regulation: An Initial Survey of US Environmental, Health, and Safety Regulation' (2010) 77 Technological Forecasting and Social Change 951.

Sub-Recommendation 1.1 Sunset Clauses – Any future statutory requirements which require or encourage the use of algorithmic systems in criminal justice should be subject to sunset clauses requiring their automatic, full qualitative review.

8.2 Compliance Capacity

Lawfulness of algorithmic systems might be a difficult challenge for some bodies, as the expertise needed to keep systems in line with relevant regimes such as human rights, anti-discrimination and data protection can be difficult to obtain and retain – especially when it has a technical component which is much in demand at the moment. It then becomes useful to add certain consolidating 'touchpoints' to ensure organisations procuring algorithmic tools in the criminal justice system have adhered to the rules.

In general, as a high-level recommendation, the Commission recommends that oversight of compliance must be improved across the board.

Recommendation 1 Oversight – A range of new mechanisms and institutional arrangements should be created and enhanced to improve oversight of algorithms in the criminal justice system.

One challenge the Commission took note of concerned the difficulty of rigorous oversight in public-private relations around algorithmic systems. Around the world, there are examples of important decisions about values, including in criminal justice, being outsourced to private entities. There are different political views on the suitable role of private entities in carrying out criminal justice-related duties, which are not for the Commission to examine, but there is one common foundation which should not be touched: value-laden decisions around the design of systems in criminal justice should never be outsourced.

Sub-Recommendation 4.1 Value-laden Decisions and Outsourcing – Value-laden decisions, such as problem definition, structuring, or choice between trade-offs in models, should never be explicitly or implicitly outsourced, for example through contracting or procurement.

To ensure this requires careful management of the design and deployment of computing within a criminal justice context. One important touchpoint can be found in the procurement of systems. At the procurement stage, many decisions are made which are practically hard to alter downstream. Opaque systems, or systems not able to be easily technically audited by the bodies responsible for them, make lawfulness downstream substantively and procedurally hard to achieve. The Commission therefore suggests that the Government develop a statutory procurement code for algorithmic systems in criminal justice, alongside a duty linked to this code.

Sub-Recommendation 4.3 Statutory Procurement Code – A procurement code for algorithmic systems in criminal justice should be developed, and a duty for relevant actors to adhere to it made a binding statutory requirement with a credible enforcement mechanism.

To address the lack of sociotechnical expertise in bodies dealing with algorithmic systems, it must be rigorously assessed and reported on. One ideal body for this is the Government's new Centre for Data Ethics and Innovation, designed to steward the wider data landscape for the public good. The body is not, however, on a statutory footing, which limits its independence and longevity. The Commission recommends it be placed upon one as soon as possible, with a requirement – among other responsibilities – to examine and report on the capacity of public bodies, including those in criminal justice, to grapple with the algorithmic issues in this report and beyond.

Sub-Recommendation 1.4 Centre for Data Ethics and Innovation – The Centre for Data Ethics and Innovation should be given statutory footing as an independent, parliamentary body, with a statutory responsibility for examining and reporting on the capacity for public bodies, including those in criminal justice, to analyse and address emerging challenges around data and society in their work, and develop a taxonomy of concepts important to algorithmic systems across sectors and domains.

Recommendation 6 Analytical Capacity and Capability – Significant investment must be carried out to support the ability of public bodies to understand the appropriateness of algorithmic systems, and where appropriate, how to deploy them responsibly.

The main regulator for many issues around data is, and remains, the Information Commissioner. The role of the Information Commissioner has greatly expanded, with the same speed as the digital age, but despite some investment, her resources have not kept pace. The Commission is concerned that the Information Commissioner needs a step-change in resources to appropriately govern algorithms in the criminal justice system in relation to the responsibilities described earlier, and makes a recommendation to that effect.

Sub-Recommendation 1.2 Capacity of the Information Commissioner – The Information Commissioner must be adequately resourced to examine algorithmic systems with rigour on a proactive, rather than predominantly reactive basis.

One of the first tasks for an Information Commissioner with an increased capacity to consider issues in criminal justice should be to create a suitable code of practice. The Law Society stands ready to engage as a key stakeholder in this process, and the interest in this Commission has strongly indicated a range of other interested actors would not hesitate to participate either. Some existing work exists, such as the ALGO-CARE framework proposed by law academic Marion Oswald alongside colleagues from other universities and police forces.[280] This framework considers issues connected to the headings of Advisory, Lawful, Granularity, Ownership, Challengeable, Accuracy, Responsible and Explainable.

There are different ways for the Information Commissioner to create codes of conduct or codes of practice, and each have drawbacks and benefits. The GDPR itself has explicit provisions for developing codes of conduct, which have been highlighted as potent governance tools for algorithmic systems.[281] The codes of conduct under the GDPR[282] can

[280] Oswald M, Grace J, Urwin S, and Barnes GC, 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and "Experimental" Proportionality' (2018) 27 Information & Communications Technology Law 223.

[281] Edwards L and Veale M, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18.

[282] GDPR, art 40.

be adopted alone by a regulator where they relate only to national matters (as opposed to going through the EDPB); however, they only relate to issues within the scope of the GDPR, whereas many criminal justice algorithms in practice are likely to fall in whole or in part under the provisions of the Law Enforcement Directive and the Data Protection Act 2018 Part 3.

The Information Commissioner can make guidance for the purposes of explaining and communicating the regulation, but this guidance does not have a basis or requirement in law. Instead, the Data Protection Act 2018 contains several requirements for the Information Commissioner to prepare codes of practice on issues including data sharing and age-appropriate design. Failure to adhere to these codes of practice does not of itself make a person liable to legal proceedings in a court or tribunal; however, the codes are admissible as evidence, and must be taken into account by a court, tribunal or Information Commissioner in the execution of her functions if applicable and considered as relevant.[283]

The Secretary of State can oblige the Information Commissioner to prepare further codes – which have relevance to all data protection legislation, including Part 3 of the Data Protection Act 2018 – and it is this power the Commission suggests should be used in the context of algorithms in the criminal justice system.

Code of Practice for Algorithmic Systems in Criminal Justice – The Government should request and resource the Information Commissioner to create a code of practice for algorithmic systems in criminal justice under the Data Protection Act 2018 s128(1).

Similarly, a code of practice is required in relation to the operation of freedom of information legislation to algorithmic systems and processes, including in the criminal justice domain. Codes of practice are envisaged under the Freedom of Information Act 2000, where they may be issued by the Minister for the Cabinet Office,[284] and guidance more generally can be issued by the Information Commissioner.[285]

Sub-Recommendation 3.3 Information Rights around Algorithmic Systems – The Government and/or Information Commissioner should provide guidance on how Freedom of Information Rights apply to value-laden, algorithmic software systems, particularly in the criminal justice sector.

8.3 Transparency, Explanation and Justification

The right to a fair trial also has elements of requiring reasoning of decisions, the exact requirements of which will require consideration of the nature of the decision and the circumstances of the case.[286] Where algorithmic systems are used in the context of a trial, it seems clear that there will be an expectation of explanation and justification capacity.[287]

As a result, it seems important that the ability to oversee these systems is ensured from the start. The UK Government's Data Ethics Framework emphasises the importance of

[283] Data Protection Act 2018 s 127.

[284] Freedom of Information Act 2000 s 45.

[285] Freedom of Information Act 2000 s 47(2).

[286] *García Ruiz v. Spain*, App no 30544/96 (ECtHR, 12 Jan 1999).

[287] See generally 'Reasoning of judicial decisions' in ECtHR, *Guide on Article 6 of the European Convention on Human Rights (criminal limb)* (30 April 2019).

considering issues such as explainability upstream in procurement and early development,[288] and this should be echoed in other sectors, such as justice.

Dr Reuben Binns and colleagues note that explanation quality – or ‘interpretability’ – does not have a formal definition or a standard evaluation methodology shared amongst machine learning researchers. They concluded that careful consideration must be paid to the ways in which such information is provided, and that “as lawmakers legislate for mandatory provision of information to decision-subjects, human-computer interaction research has much to offer in how such information should be extracted, presented and delivered”.[289]

A number of witnesses to the Commission noted that while some degree of transparency was desirable, context was important. Professor Roger Brownsword argued that “to say there must be transparency is just the beginning; there need to be debates about how much transparency and in which contexts”. Professor Lilian Edwards warned of a “transparency fallacy to rival the notice-and-consent fallacy”, where assuming that individual explanations were helpful in regulating an entire value-laden system with skewed power dynamics was problematic. From a different standpoint, barrister Matthew Lavy expressed concern that technologies could be gamed or fooled with higher propensity if they adopted explanation facilities, and that “if all components of decision-making in the justice system must be susceptible to explanation, it follows that the vast majority of current-generation machine learning technologies (whether recurrent neural networks, support vector machines or other exotica) must be ruled out”.

Other evidence to the Commission suggested that data protection transparency provisions – currently available to individual data subjects – should have a more open public basis to allow for public scrutiny and review (Roger Bickerstaff) and that processing of data about groups is not subject to access rights which points to the need to recognise “some type of group privacy right in data protection law” (Professor Edwards).

The provision of accessible information is a key part of human rights treaties themselves. For instance, article 5 (1)(c) of the ECHR requires the existence of reasonable suspicion before an arrest is lawful. Reasonable suspicion requires “the existence of facts or information which would satisfy an objective observer[290],” and the basis of any detention must be subject a prompt and independent verification by a judicial officer (article 5(3)). A key provision of the same article (article 5(2) requires the person arrested to be “informed promptly, in a language he understands, of the reason for his arrest…”.

The right to a remedy implies the right to a reasoned and individualised decision. An individual should be able to enquire, for instance, as to why they have been placed on a ‘Gangs Matrix’[291], or why they have been identified as a potential suspect by predictive software, or why they have been denied access to a rehabilitation programme.

Consequently, the Commission recommends explanation facilities for algorithmic systems in criminal justice at two levels: the individual and the societal.

[288] DCMS, *Data Ethics Framework* (HM Government 2018).

[289] Binns R, Van Kleek M, Veale M, Lyngs U, Zhao J, and Shadbolt N, ‘“It’s Reducing a Human Being to a Percentage”; Perceptions of Justice in Algorithmic Decisions’ Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI 2018) (ACM 2018).

[290] Fox, Campbell and Hartley v UK, (1990) para 32

[291] Launched by the Metropolitan Police in 2012, the Gangs Matrix is a database of suspected gang members in London, and has been criticised as being racially discriminatory, see Amnesty International T’rapped in the Matrix: Secrecy, stigma, and bias in the Met’s Gangs Database’ – May 2018

Sub-Recommendation 4.4 Individual Explanation Facilities and Remedies – Algorithmic systems in criminal justice must have explanation facilities focused on each decision or measure, designed to help individuals and users assess whether a given output is justified, and whether they should seek a remedy through the courts

Sub-Recommendation 4.5 Societal Explanation Facilities – Algorithmic systems in criminal justice must have explanation facilities designed to allow broader internal and external scrutiny, such as over the general logics, functioning, behaviour and impact of the models.

These explanations are not the only form of relevant transparency, however. As this report has demonstrated, understanding the extent of the deployment of algorithmic systems in the criminal justice sector is extremely challenging, relying on those actors willing to give evidence or brief the media; freedom of information law; or other sources of fortuitous or serendipitous release. This is inappropriate for a topic with such risks for individuals' rights and freedoms. The Commission proposes two main approaches to this challenge.

Sub-Recommendation 1.7 National Register of Algorithmic Systems – A register of algorithmic systems in criminal justice should be created, including those not using personal data, alongside standardised metadata concerning both their characteristics, such as transparency and discrimination audits and relevant standard operating procedures, and the datasets used to train and test them. Leadership of this could be taken by the Centre for Data Ethics and Innovation, as the Centre matures, in an open consultation procedure considering the criteria and thresholds for systems included in this register.

A national register would allow clear inroads for the Commission's second recommendation, which covers more detailed and substantive oversight.

Sub-Recommendation 1.6 Public Interest Access – A facility should be established to enable secure access to algorithmic systems in use by or on behalf of public bodies in the criminal justice system for researcher and journalistic oversight. The British Library and the Centre for Data Ethics and Innovation could be candidates for coordinating this effort.

Organisations using this transparency need an ability to act, and flag misuses of systems. They may not always be able to take a legal case, due to lack of funds or insufficient interest in the case to qualify for judicial review, for example. A neat solution exists in the GDPR, however, at Article 80(2). Article 80(2) resembles, to some degree, the super-complaint mechanism found in many existing UK laws.

For example, the Enterprise Act 2002, s11 enables a consumer body designated by the Secretary of State to make a complaint to the Competitions and Markets Authority (or a selection of other authorities,[292] "that any feature, or combination of features, of a market in the United Kingdom for goods or services is or appears to be significantly harming the interests of consumers". The Financial Services and Markets Act 2000, s234C, provides that a "designated consumer body may make a complaint to the [Financial Conduct Authority] that a feature, or combination of features, of a market in the United Kingdom for financial

[292] The Enterprise Act 2002 (Super-complaints to Regulators) Order 2003, SI 2003/1368.

services is, or appears to be, significantly damaging the interests of consumers". The Financial Services (Banking Reform) Act 2013, s68, also provides a "designated representative body may make a complaint to the Payment Systems Regulator that a feature, or combination of features, of a market in the United Kingdom for services provided by payment systems is, or appears to be, significantly damaging the interests of those who use, or are likely to use, those services".

Article 80(2) would allow a properly constituted body to exercise some data protection rights, such as the right to complain to a supervisory authority or to seek a judicial remedy, without requiring a data subject. The Government did not implement Article 80(2), which was optional for Member States, but instead committed to review a range of collective proceedings possible or not yet implemented under the GDPR 30 months from the passing of the Data Protection Act 2018.[293] This would be a prime opportunity for a new mechanism of oversight to be created.

Sub-Recommendation 1.5 Super-complaints – The Government should make provisions for Article 80(2) of the GDPR, which allows civil society organisations to complain to the ICO and seek a judicial remedy on behalf of a group rather than an individual. This provision should apply to the whole Data Protection Act 2018, including Part 3, rather than just the GDPR.

8.4 Discrimination and Computational Power

Algorithmic systems both highlight and exacerbate issues of discrimination and power imbalances. There are many reasons for this, which mirror the myriad reasons that bias and powerlessness manifest in other parts of society, and as a result, approaching this issue requires decision makers to zoom out, and see the challenge holistically, free from technological determinism or solutionism.

Algorithmic systems might be useful in understanding and monitoring discrimination within a criminal justice context. Increased digitisation of information within a justice context might allow for monitoring and evaluation, similarly to the recent gender pay gap reporting rules in the UK.[294] Such insight might be beneficial to understanding barriers to access to justice, for example. The Commission did not yet observe this promise in practice, although several witnesses raised the possibility of algorithms being used to provide analysis and oversight of fairness issues. The Commission believes this is an area worth exploring further, and recommends research projects and educational capacity-building be undertaken between public bodies and universities to pilot these efforts.

Sub-Recommendation 6.2 Research Support – The Government should support joint research projects between universities and actors in the justice sector around applied algorithmic systems, including how algorithmic analysis can promote equity in and access to justice.

[293] Data Protection Act 2018, s 189.
[294] The Equality Act 2010 (Gender Pay Gap Information) Regulations 2017.

Sub-Recommendation 6.3 Training Support – The Government should support universities in offering educational programmes for public interest practitioners sitting at the intersection of technology, law and human rights.

Most issues around algorithmic systems and discrimination that the Commission encountered did, however, centre on the ability for these algorithms to perpetuate injustice through relying on biased datasets or disregarding the environmental complexities and the side effects of optimisation.[295]

The Commission has several recommendations applicable broadly to systems with discriminatory potential, such as hotspot policing. Such recommendations must consider that personal data may not necessarily be at the core of these decisions, even if social issues are at stake. As a result, data protection may not govern machine learning without overstretching its original purpose.[296] This leads the Commission to make an overarching recommendation of the importance to consider all legal frameworks in a connected and holistic manner, rather than assuming that consequential algorithms in the justice system will be successfully governed by data protection legislation alone.

Recommendation 3 Protection beyond Data Protection – Existing regulations concerning fairness and transparency of activities in the justice sector should be strengthened in relation to algorithmic systems.

In this vein, the Commission calls for further clarity and emphasis on existing legal provisions, such as the public sector equality duty.[297] This duty applies to public authorities carrying out their functions, including in policing. This duty combines and extends previous equality duties – public authorities have been under a general duty to have due regard to the need to promote race equality since 2001; disability equality since 2006; and sex equality since 2007.[298]

The motivation behind the public sector equality duty (PSED) is well explained in a 2010 report published by the Government Equalities Office. In it, they state that individuals with certain protected characteristics “all have different needs and may face different levels of discrimination or barriers to accessing services”, going on to note that it “is only right that we use the powerful tool of the public sector to help eliminate any discrimination they may face”.[299]

There are two reasons the PSED is a useful tool to extend for the purposes of algorithmic governance. Firstly, it is a deep duty.

> “The [duties] apply to all "functions" performed by public authorities and, importantly, they thus apply not only to discretionary decisions with which administrative law has historically been concerned, but they also reach right down to day-to-day operational decisions. They thus permeate deep into public service provision and public administration. The duties apply to the

[295] Overdorf R, Kulynych B, Balsa E, Troncoso C, and Gürses S, ‘POTs: Protective Optimization Technologies’ [2018] arXiv:180602711 [cs].

[296] See generally on this overstretch Veale M, Binns R, and Edwards L, ‘Algorithms that Remember: Model Inversion Attacks and Data Protection Law’ (2018) 376 Phil Trans R Soc A 20180083.

[297] Equality Act 2019, s 149.

[298] Sales J, ‘The Public Sector Equality Duty’ (2011) 16 Judicial Review 1.

[299] Government Equalities Office, Equality Bill: Making it work. Policy proposals for specific duties (HM Government 2010).

> daily activities of public servants such as police officers, social workers and teachers. A police officer walking his or her beat must do so with "due regard" to the identified "needs".”[300]

Insofar as algorithms reach into the day-to-day decisions of front-line workers in parts of the criminal justice system covered, such as the police, it is a deep tool with far-reaching consequences.

Secondly, the PSED also applies to procurement processes and decisions. Contractors will need to meet requirements of the PSED.[301] As we are discussing technical tools with the potential for discrimination, considering the PSED as a way of enforcing algorithmic bias analysis in general in procurement seems a promising avenue to utilise and potentially extend existing legal frameworks and provisions.

A popular way to demonstrate PSED compliance has been through undertaking an equality impact assessment (EqIA)[302], although political concerns about bureaucratic load mean these are somewhat less commonly found today. An EqIA is a loose and variable tool that is used to consciously evaluate policies or interventions and anticipate their equality-related effects. A range of guidance documents exist to guide policy makers in the creation of these documents, which broadly recommend explicitly considering and discussion policy purpose, stakeholder participation, analytic and ongoing monitoring approaches, and any proposed efforts to alleviate negative impact.[303] PSED principles established in *Brown v Secretary of State for Work and Pensions*[304] indicate that the duty must be fulfilled both before and during the time when a public authority is considering a policy, consciously rather than justified ex post facto, and that it is good practice to keep records on PSED compliance. This has made an EqIA a preferred and understood option, although certainly not the only way of demonstrating compliance. A non-statutory UK Government requirement mandating an EqIA for every policy was scrapped in November 2012 by the then-Government; however, in this context it is reasonable to determine that the degree of uncertainty, and the risk of long-term harm to the democratic legitimacy and trust in the criminal justice system, on balance warrant a responsible and transparent response.

The Commission feels that EqIAs in connection with the PSED are a powerful tool for algorithmic systems in the criminal justice sector and similar algorithmic systems in the public sector more broadly, and are a good case for mandating the use of such processes under existing law where bodies fall under this responsibility.

Sub-Recommendation 3.1 Public Sector Equality Duty – Given the importance of countering discrimination within algorithmic systems, Equality Impact Assessments should be formalised as a requirement before deploying any consequential algorithmic system in the public sector and these should be made proactively, publicly available.

The PSED is no panacea, however. One challenge with the provision is that it is accessible to individuals primarily through the mechanism of judicial review. Although a route which can

[300] Hickman T, ‘Too hot, too cold or just right? The development of the public sector equality duties in administrative law’ 2013 Public Law 325.

[301] Cabinet Office, *Procurement Policy Note –Public Procurement and the Public Sector Equality Duty* (Information Note 01/1328 January 2013, HM Government 2013).

[302] *An equality impact assessment (EqIA) is a process designed to ensure that a policy, project or scheme does not discriminate against any disadvantaged or vulnerable people.*

[303] Equality and Human Rights Commission, Public Sector Equality Duties and financial decisions – a note for decision makers (2017).

[304] Brown v. Secretary of State for Work and Pensions EWHC 3158 (Admin) 2008.

be pursued, judicial review is burdensome, particularly given the need to demonstrate interest in the matter to be reviewed, given that many algorithmic systems contribute to cumulative disadvantage rather than a specific incident likely to result in an individual wishing to pursue a legal remedy.

A further challenge is the lack of necessary connection between the protected categories in equality law and the potential harms we see from algorithmic systems. Many of these harms appear to be concerning on other, more difficult to determine lines, such as socioeconomic divides.[305] Luckily, there are similar provision in UK law which can be repurposed towards these ends. The socioeconomic equality duty exists as s1 of the Equality Act 2010, although it has only been commenced in Scotland. The Commission recommends commencing this section across the whole of the UK, particularly with regards to algorithmic systems in criminal justice.

Sub-Recommendation 3.2 Socioeconomic Equality Duty – Given algorithmic systems' high potential for socioeconomic discrimination, the Government should commence the socioeconomic equality duty in the Equality Act 2010 s1 in England and Wales, at least with regard to algorithmic decision-support systems.

In general, challenges with these duties and Equality Impact Assessment more broadly surround the substantive and resource-intensive analysis required to sufficiently understand the role of a system. This is difficult to undertake in low capacity environments, as analysis of the role of an algorithmic system not only requires significant statistical mastery, but also an ability to analyse the entire system as a whole, measuring concepts like non-discrimination across a system more widely, and understanding the downstream impacts of the system.[306] In section 8.2, we discussed the need for increased capacity for compliance and responsibility more broadly. This is especially required to ensure that impact assessments around algorithms do not, as Commission witnesses Professor Lilian Edwards noted they:

> "…bring with them a real danger of formalistic bureaucratic overkill alongside a lack of substantive change: a happy vision for more form-filling jobs and ticked boxes, but a sad one for a world where automated algorithms do their jobs quietly without imperilling human rights and freedoms, especially privacy and autonomy."[307]

8.5 Concluding remarks

In this report, the Commission has found a general and concerning lack of openness or transparency about the use of algorithmic systems in criminal justice across England and Wales. This was concerning, as the high-stakes decisions and measures taken in the justice system demand extremely careful deployment. There are significant challenges of bias and discrimination, opacity and due process, consistency, amenability to scrutiny, effectiveness, disregard of qualitative and contextual factors, against a backdrop of the potential of these systems to more deeply change the nature of the evolution of the law.

[305] See generally Eubanks V, *Automating Inequality* (St Martin's Press 2018).
[306] Selbst AD, Boyd D, Friedler SA, Venkatasubramanian S, and Vertesi J, 'Fairness and Abstraction in Sociotechnical Systems' in Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19, New York, NY, USA, ACM 2019).
[307] See generally Edwards L and Veale M, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18, 80.

It appears to the Commission to be a myth that the technologies being deployed are so technically novel that they cannot be critically assessed by multi-disciplinary teams for their effectiveness, their conformity to real challenges, and their potential for unintended and undesirable side-effects, particularly from optimising for some goals or aspects of an issue to the detriment of others. In-house analytical capacity to analyse, oversee and maintain these systems is, in many instances, lacking, but key. Still, in-house capacity is only one piece of the puzzle. Governing algorithmic systems in criminal justice usually brings multi-dimensional tensions and value-laden choices to grapple. These tensions emerge at many different points in development, deployment and maintenance, and are usually not between a ‘bad’ and a ‘good’ outcome, but between different values that are societally held to be of similar importance. It insufficient and unacceptable for the bodies and agencies involved to make these decisions alone, requiring instead the engagement of broad stakeholders including civil society, academia, technology firms and the justice system more broadly.

Risks of systems being gamed is real, but often overstated in relation to the risks from lack of openness, engagement, and the loss of trust in procedural justice and the rule of law. Such risks stem especially from what are effectively policy decisions baked into algorithmic systems being made invisibly and unaccountably by contractors and vendors. Political design choices, especially those in criminal justice, should never be outsourced.

In the course of evidence-taking, Commission became heavily concerned that some systems and databases operating today, such as facial recognition in policing or some uses of mobile device extraction, lack a clear and explicit lawful basis, as well as unclear proven algorithmic performance. This must be urgently examined, publicly clarified and rectified if necessary. While the United Kingdom has more explicit provisions covering algorithmic systems than many other parts of the world, these contains significant omissions and loopholes that need joined-up consideration. Several clarifications and changes to data protection legislation, procurement codes, freedom of information law, equality duties and statutory oversight and scrutiny bodies have been recommended in this report. These would provide key safeguards to the integrity of criminal justice in the digital age.

Many of the heavily individualised, legal safeguards proposed to algorithmic systems in commercial domains, such as individual explanation rights, are unlikely to be very helpful in criminal justice, where imbalances of power can be extreme and are exacerbated by the dwindling availability of legal aid. Societal, systemic oversight must be placed at the forefront of algorithmic systems in this sector, which will require innovative and world-leading policies.

The Commission believes that the United Kingdom has a window of opportunity to become a beacon for a justice system trusted to use technology well, with a social licence to operate and in line with the values and human rights underpinning criminal justice. It must take proactive steps to seize that window now.

9 Annexes

This section contains supplementary information around the functioning of the Commission.

9.1 Engaged Stakeholders

We are grateful to all those who participated in this investigation. The following gave oral evidence, submitted written evidence or participated in an in-depth interview.

Dr Reuben Binns, University of Oxford and ICO
Dr Nikos Aletras, Sheffield University
Dr Michael Veale, UCL/Alan Turing Institute
Professor Burkhard Schafer, University of Edinburgh
Professor Lilian Edwards, Newcastle University
Dr Ricardo Silva, University College London
Professor Lorna McGregor, Essex University
Alexander Babuta, RUSI
Roger Bickerstaff, Bird & Bird
Benoit Van Asbroek, Bird & Bird
Professor David Hand OBE
Chief Constable Michael Barton, Durham Constabulary
Marion Oswald, Winchester University
Matthew Lavy, Barrister, 4 Pump Court
Professor Karen Yeung, University of Birmingham
Dr Adrian Weller, U of Cambridge/Alan Turing Institute
Simon Burall, Involve
Ed Bird , Solomonic
Gideon Cohen, Solomonic
Sue Daley, techUK
Guy Cohen, Privitar
Dr Hannah Knox, University College London
Alesis Novik, AimBrain
Nikita Malik, Henry Jackson Society
Dr Vicky Kemp, University of Nottingham
Lord Clement Jones, Chair of the House of Lords Select Committee on AI & DLA Piper
Professor Richard Susskind OBE
Jamie Susskind, Littleton Chambers
Alvin Carpio, Fourth Group
Dr Vicky Kemp, Nottingham University
Dr Hannah Knox, University College London
Silkie Carlo, Big Brother Watch
Jacob Turner, Barrister, Fountain Court Chambers
Peter Wells, Open Data Institute
David Powell, Hampshire Police
Judith Jones, Information Commissioner's Office
Hannah Couchman, Liberty
Dr Jiri Novak, Chair of the IT Law Committee of the CCBE
Professor William Wong, Middlesex University
Catherine Miller, doteveryone
Clementina Barbaro, Council of Europe
Stephane Leyenberger, Council of Europe
Sharan Johnstone, University of South Wales
Professor Martin Innes, Cardiff University
Mike Edwards, University of South Wales
Dr Adam Wyner, Swansea University
Adam Curtis, Hoowla
Dr Bernadette Rainey, Cardiff University
Karl Foster, Blake Morgan
Emma Erskine-Fox, TLT
Inspector Scott Lloyd, South Wales Police
Paris Theodorou, Hodge, Jones & Allen
Emma Wright, Kemp Little
Richard Goodwin, HMCTS
Toby Unwin, Premonition
James Chandler, Benevolent AI
Professor Roger Brownsword, King's College London
Sam Spivak, Kira Systems
Dr Sandra Wachter, University of Oxford
Huw Bowden, Bowden Jones
Mark Blake, BTEG
Dr Stephen Castell, Castell Consulting
Richard Pinch, IMA
Louise Waltham, The Royal Statistical Society
Stephanie Balsys, Mischon de Reya
Fraser Matcham, Legal Utopia Limited
Julian Sole, Fleishman Hillard Fishburn
Stephen Mason, Barrister
Noel Corriveau, Treasury Board, Government of Canada
Julien Pelletier-David, Barreau du Quebec
Charles Kerrigan, CMS
Chris de Silva, NEC Corporation

9.2 Lines of Enquiry

What AI-based technologies are **currently in use** in the justice system in England and Wales? Or internationally?

What AI-based technologies **are in development** which may have an application in the justice system?

What are some of the **benefits** that can be derived from the use of AI and other emerging technologies based on machine learning in the justice system?

What are some of the **dangers?**

What can we **rely on** the technologies to do well? And what should we **not rely upon** them to do?

What lessons can be drawn from the application of these tools in **other domains?**

How has **industry reacted** to the fears that some have identified in the uncontrolled use of AI in fields such as justice, or social policy?

What **responsibilities** do developers have? What responsibility do suppliers have? What responsibility do the users of these technologies have? And where does accountability rest?

How does the **profit motive** effect decision making on issues such as ethics? Transparency?

What role does industry feel **governments** have in these issues?

What measures have **industry** collectively developed? How can the effectiveness of these responses be measured?

How does industry seek **to engage** with a wider stakeholder group? Should it? And can it do so effectively?

What are the **constraints on development** which delivers more safeguards?

What are **the implications for the Rule of Law** and Fundamental Freedoms from using algorithms in the justice system?

What **role do citizens** have in setting future norms? How can this be effective? Where does responsibility lie for such engagement?

What **paradoxes exist** in the debate between agency, privacy, regulation, innovation, speed and efficiency and safety?

What do we know about **human behaviour –** from disciplines such as anthropology and political science – to understand the rationale for the trade-offs and values of today and the impacts on longer term understandings of the social contract?

Are these issues **local, domestic, regional, global**? How would business prefer this to be handled?

What can we learn from regional exploration in this area about the likelihood of e.g. **Europe-wide consensus**?

What are the challenges to achieving consensus?

What does governance – agile governance possibly – look like?

References

Abdul A, Vermeulen J, Wang D, Lim BY, and Kankanhalli M, 'Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda' in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18, ACM 2018)

Aecus European Innovation Awards, "Risk-based" predictive modelling to assist in the tackling of Burglary crime within Force Area (2015)

AI Now Institute, Litigating Algorithms: Challenging Government use of Algorithmic Decision Systems (New York University 2018).

Ajunwa I, Crawford K, and Schultz J, 'Limitless Worker Surveillance' (2017) 105 Calif L Rev 735.

Alston P, 'Statement on Visit to the United Kingdom, by Professor Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights' (United Nations 2018)

Amankwaa AO and McCartney C, 'The UK National DNA Database: Implementation of the Protection of Freedoms Act 2012' (2018) 284 Forensic Science International 117

Amnesty International UK, Trapped in the Matrix: Secrecy, stigma, and bias in the Met's Gangs Database (2018)

Azavea, HunchLab: Under The Hood (Azavea 2015)

Barocas S and Selbst AD, 'Big Data's Disparate Impact' (2016) 104 California Law Review 671

BBC News, 'Rape victims among those to be asked to hand phones to police' (BBC 29 April 2019)

Big Brother Watch, Face Off: The Lawless Growth of Facial Recognition in UK Policing (Big Brother Watch 2018)

Binns R, 'Data protection impact assessments: A meta-regulatory approach' (2017) 7(1) International Data Privacy Law 22 doi:10/cvct.

Binns R, Van Kleek M, Veale M, Lyngs U, Zhao J, and Shadbolt N, '"It's Reducing a Human Being to a Percentage"; Perceptions of Justice in Algorithmic Decisions' Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI 2018) (ACM 2018).

Biometrics and Forensics Ethics Group Facial Recognition Working Group, Ethical Issues Arising from the Police Use of Live Facial Recognition Technology (Biometrics and Forensics Ethics Group 2019)

Black J, 'The Emergence of Risk-Based Regulation and the New Public Risk Management in the United Kingdom' (2005) 3 Public Law 512

Bowers KJ, Johnson SD and Pease K, 'Prospective Hot-Spotting: The Future of Crime Mapping?' (2004) 44 British Journal of Criminology 641

Breiman L, 'Random Forests' (2001) 45 Machine Learning 5

Brundage M and others, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (2018) https://maliciousaireport.com/.

Buolamwini J and Gebru T, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' in Conference on Fairness, Accountability and Transparency (FAT* 2018) (2018)

Burrell J, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 205395171562251

Bygrave LA, 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling' (2001) 17 Computer Law & Security Review 17

Cabinet Office, Procurement Policy Note –Public Procurement and the Public Sector Equality Duty (Information Note 01/1328 January 2013, HM Government 2013

Carton S, Helsby J, Joseph K, Mahmud A, Park Y, Walsh J, Cody C, Patterson CE, Haynes L, and Ghani R, 'Identifying Police Officers at Risk of Adverse Events' in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '16, New York, NY, USA, ACM 2016)

Centre for Legal Research, Annual Report 2013/2014 (University of West of England 2014)

Chouldechova A, 'Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments' (2017) 5 Big Data 153

Citron DK, 'Technological Due Process' (2008) 85 Wash U. L. Rev. 1249

Coglianese C and Lehr D, ‘Regulating by Robot: Administrative Decision Making in the Machine-Learning Era’ (2017) 105 Geo. L.J. 1147

Cohen JE, “Examined Lives: Informational Privacy and the Subject as Object” (2000) 52 Stan. L. Rev. 1373

College of Policing, The Effects of Hot-Spot Policing on Crime: What Works Briefing (College of Policing, September 2013)

College of Policing, *Body-Worn Video* (2014)

Colquitt JA, Conlon DE, Wesson MJ, Porter COLH, and Ng KY, ‘Justice at the Millennium: A Meta-Analytic Review of 25 Years of Organizational Justice Research.’ (2001) 86 Journal of Applied Psychology 425

Copas J and Marshall P, ‘The Offender Group Reconviction Scale: A Statistical Reconviction Score for Use by Probation Officers’ (2002) 47 Journal of the Royal Statistical Society: Series C (Applied Statistics) 159

Cormen TH, Leiserson CE, Rivest RL and Stein S, *Introduction to Algorithms* (MIT Press 2009)

Couchman H, ‘Policing by Machine’ (Liberty 2018)

Coutts P, Turning Point: The Police’s Production and Use of Evidence to Reduce Reoffending (Alliance for Useful Evidence, January 2018

Davies B, Innes M, and Dawson A, An Evaluation of South Wales Police’s Use of Automated Facial Recognition (Universities’ Police Science Institute and Crime & Security Research Institute, Cardiff University 2018)

De Wetenschappelijke Raad voor het Regeringsbeleid (WRR) [Dutch Scientific Council for Government Policy], Big Data in Een Vrije En Veilige Samenleving [Big Data in a Free and Safe Society] (Amsterdam University Press 2016)

Debidin M (ed), A Compendium of Research and Analysis on the Offender Assessment System (OASys) 2006-2009 (Ministry of Justice 2009)

Delacroix S, ‘Taking Turing by Surprise? Designing Digital Computers for Morally-Loaded Contexts’ [2018] arXiv:180304548

Delacroix S, ‘Computer Systems Fit for the Legal Profession?’ (2018) Legal Ethics, doi:10.1080/1460728x.2018.1551702

Delacroix S and Veale M, ‘Smart Technologies and Our Sense of Self: Going Beyond Epistemic Counter-Profiling’ in K O’Hara and M Hildebrandt (eds), Law and Life in the Era of Data-Driven Agency (Edward Elgar 2019) doi:10/gfzvz9

Deloitte and Reform, The State of the State 2018-19 (2019)

Dencik L, Hintz A, Redden J, and Warne H, ‘Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services’ (Data Justice Lab, Cardiff University 2018)

Denham E, ‘Facial Recognition Technology and Law Enforcement’, The Information Commissioner’s Office Blog (14 August 2018)

Desmarais SL, Johnson KL, and Singh JP, ‘Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings’ in Handbook of Recidivism Risk/Needs Assessment Tools (John Wiley & Sons, Ltd 2017)

Domingos P, ‘A Few Useful Things to Know about Machine Learning’ (2012) 55 Communications of the ACM 78.

Dunleavy P, Margetts H, Bastow S, and Tinkler J, Digital Era Governance (Oxford University Press 2006)

Dus A, Nelle D, Stock G and Wagner GG (eds), Facing the Future: European Research Infrastructures for the Humanities and Social Sciences (SCIVERO Verlag 2014)

ECtHR, *Guide on Article 6 of the European Convention on Human Rights (criminal limb)* (30 April 2019).

Edwards L and Urquhart L, ‘Privacy in Public Spaces: What Expectations of Privacy Do We Have in Social Media Intelligence?’ (2016) 24 International Journal of Law and Information Technology 279

Edwards L, Martin L, and Henderson T, ‘Employee Surveillance: The Road to Surveillance is Paved with Good Intentions’ (SSRN Scholarly Paper, Social Science Research Network 18 August 2018)

Edwards L and Veale M, ‘Slave to the Algorithm? Why a “Right to an Explanation” Is Probably Not the Remedy You Are Looking For’ (2017) 16 Duke Law & Technology Review 18

Edwards L and Veale M, ‘Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?’ (2018) 16 IEEE Security & Privacy 46

Elliott M, ‘Has the Common Law Duty to Give Reasons Come of Age Yet?’ (University of Cambridge Faculty of Law Research Paper No. 7/2012, University of Cambridge 2012)

Ensign D, Friedler SA, Neville S, Scheidegger C, and Venkatasubramanian S, ‘Runaway Feedback Loops in Predictive Policing’ in Conference on Fairness, Accountability and Transparency (FAT* 2017) (PMLR 2018)

Equality and Human Rights Commission, Public Sector Equality Duties and financial decisions – a note for decision makers (2017)

Eubanks V, *Automating Inequality* (St Martin’s Press 2018),

Freed D, Palmer J, Minchala D, Levy K, Ristenpart T, and Dell N, ‘A Stalker’s Paradise: How Intimate Partner Abusers Exploit Technology’ in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems–CHI ’18 (ACM Press 2018)

Gandy Jr OH, ‘Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems’ (2010) 12 Ethics and Information Technology 29

Gandy Jr OH, Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage (Routledge 2009)

Government Equalities Office, Equality Bill: Making it work. Policy proposals for specific duties (HM Government 2010)

Gürses S and Hoboken J van, ‘Privacy after the Agile Turn’ in E Selinger, J Polonetsky, and O Tene (eds), *The Cambridge Handbook of Consumer Privacy* (1st edn, Cambridge University Press 2018)

Hickman T, ‘Too hot, too cold or just right? The development of the public sector equality duties in administrative law’ 2013 Public Law 325

Hildebrandt M, Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology (Edward Elgar 2015).

Hill R, ‘ICO to Probe Facial Recog amid Concerns UK Cops Can’t Shake Their Love for Unregulated Creepy Tech’, The Register (3 December 2018)

HM Government, National.Cyber Security Strategy 2016 to 2021 (HM Government 2016).

HM Crown Prosecution Service Inspectorate and HM Inspectorate of Constabulary, Delivering Justice in a Digital Age. (Criminal Justice Joint Inspection 2016)

HM Government, The HMCTS reform programme (GOV.UK 2019) https://www.gov.uk/guidance/the-hmcts-reform-programme

Hood C, ‘Public Service Management by Numbers: Why Does It Vary? Where Has It Come From? What Are the Gaps and the Puzzles?’ (2007) 27 Public Money and Management 95.

Hoogden RH van den, ‘E-Justice, Beginselen van Behoorlijke Elektronische Rechtspraak’ (PhD, Universiteit Utrecht 2007).

House of Commons Science and Technology Select Committee, Algorithms in Decision-Making, (HC 351, Fourth Report of Session 2017–19, 2018)

House of Commons Science and Technology Select Committee, Current and future uses of biometric data and technologies, Sixth Report of Session 2014-15, House of Commons: London, 25 February 2015

House of Lords Science and Technology Committee, Forensic science and the criminal justice system: a blueprint for change (HL Paper 333 2019).

Howard P, ‘Offender Group Reconviction Scale’ in Handbook of Recidivism Risk/Needs Assessment Tools (John Wiley & Sons, Ltd 2017).

Howard P, ‘The Effect of Sample Heterogeneity and Risk Categorization on Area Under the Curve Predictive Validity Metrics’ (2017) 44 Criminal Justice and Behavior 103.

Hutt O, Bowers K, Johnson S, and Davies T, ‘Data and Evidence Challenges Facing Place-Based Policing’ (2018) 41 Policing: An International Journal 339, 342–3.

Information Commissioner’s Office, Outsourcing Oversight? The Case for Reforming Access to Information Law (ICO 2019)

Information Commissioner’s Office, Guide to Law Enforcement Processing (ICO 2019)

Johnson SD, Birks DJ, McLaughlin L, Bowers KJ, and Pease K, ‘Prospective Crime Mapping in Operational Context Final Report’ (Home Office Online Report, Home Office 2007)

Johnson S, Digital Triage Devices (Cyber Kiosks) (Scottish Police Authority, 8 May 2019)

Jones ML, 'The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood' (2017) 47 Soc Stud Sci 216

Kamarinou D, Millard C and Singh J, 'Machine Learning with Personal Data' (2016) Queen Mary School of Law Legal Studies Research Paper No. 247/2016

Kaminski ME, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92 Southern California Law Review __ doi: 10/gfzx54

Kaminski ME, 'The Right to Explanation, Explained' (2019) 34 Berkeley Technology Law Journal __.

Kamiran F and Calders T, 'Classifying without Discriminating' in (IEEE February 2009) 2009 2nd International Conference on Computer, Control and Communication.

Kemper J and Kolkman D, 'Transparent to Whom? No Algorithmic Accountability without a Critical Audience' (2018) Information, Communication & Society (2018) doi:10/gfdbp6

Kershaw C, 'Interpreting Reconviction Rates' in Papers from the British Criminology Conference, Queens University, Belfast, 15–19 July 1997 (The British Criminology Conferences: Selected Proceedings, British Society of Criminology 1999) vol 2.

Kilbertus N, Gascon A, Kusner M, Veale M, Gummadi KP, and Weller A, 'Blind Justice: Fairness with Encrypted Sensitive Attributes' in J Dy and A Krause (eds), Proceedings of the 35th International Conference on Machine Learning, vol 80 (Proceedings of Machine Learning Research, PMLR 2018)

Kowalska K, Shawe-Taylor J, and Longley P, 'Data-Driven Modelling of Police Route Choice' in *Proceedings of the 23rd Conference on GIS Research UK 15th - 17th April, 2015 University of Liverpool, UK* (2015)

Kroll J, Huey J, Barocas S, Felten E, Reidenberg J, Robinson D, and Yu H, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 633

Law Commission, *Sentencing law in England and Wales: Legislation currently in force* (2015)

Leventhal GS, 'What Should Be Done with Equity Theory? New Approaches to the Study of Fairness in Social Relationships' in K Gergen, M Greenberg, and R Willis (eds), Social Exchange: Advances in Theory and Research (Plenum 1980)

Lipsky M, Street-Level Bureaucracy: Dilemmas of the Individual in Public Service (Russell Sage Foundation 2010).

London Metropolitan Police, 'Facial Recognition to Take Place in Romford' (London Met, 13 February 2019) <http://news.met.police.uk/news/facial-recogntion-to-take-place-in-romford-358589> accessed 28 April 2019

London Policing Ethics Panel, Interim Report on Live Facial Recognition (LPEP 2018)

Lum C and Koper CS, 'Evidence-Based Policing' in G Bruinsma and D Weisburd (eds), Encyclopedia of Criminology and Criminal Justice (Springer New York 2014).

Mair G, Burke L, and Taylor S, '"The Worst Tax Form You've Ever Seen"? Probation Officers' Views about OASys' (2006) 53 Probation Journal 7.

Margetts H, 'The Automated State' (1995) 10 Public Policy and Administration 88

Margetts H, Information Technology in Government: Britain and America. (Taylor and Francis 2012)

Marshall D and Thomas T, Privacy and Criminal Justice (Palgrave Macmillan 2017)

McCray LE, Oye KA, and Petersen AC, 'Planned Adaptation in Risk Regulation: An Initial Survey of US Environmental, Health, and Safety Regulation' (2010) 77 Technological Forecasting and Social Change 951

Mitchell TM, *Machine learning* (McGraw Hill 1997)

Mohler GO, Short MB, Malinowski S, Johnson M, Tita GE, Bertozzi AL, and Brantingham PJ, 'Randomized Controlled Field Trials of Predictive Policing' (2015) 110 Journal of the American Statistical Association 1399

Moore R, The Offender Assessment System (OASys): Development, validation and use in practice <https://service.mvnet.de/_php/download.php?datei_id=38021>.

Moore R (ed), A Compendium of Research and Analysis on the Offender Assessment System (Ministry of Justice Analytical Series 2015) <https://perma.cc/W2FT-NFWZ>

Nilsson P, 'How UK Police Are Using Facial Recognition Software', Financial Times (12 October 2018)

Nonaka I, 'The Knowledge-Creating Company' (1991) Nov-Dec Harvard Business Review

O'Hara K, 'The Seven Veils of Privacy' (2016) 20 IEEE Internet Computing 86

Oswald M, 'Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power' (2018) 376 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20170359.

Oswald M and others, 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and "Experimental" Proportionality' (2018) 27 Information & Communications Technology Law 223

Oswald M and Grace J, 'Intelligence, Policing and the Use of Algorithmic Analysis: A Freedom of Information-Based Study' (2016) 1 Journal of Information Rights, Policy and Practice.

Overdorf R, Kulynych B, Balsa E, Troncoso C, and Gürses S, 'POTs: Protective Optimization Technologies' [2018] arXiv:180602711 [cs].

Parrado S, 'Analytical Capacity' in Martin Lodge and Kai Wegrich (eds), The Problem-solving Capacity of the Modern State (Oxford University Press 2014)

Partnership on AI, *Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System* (2019)

Pasquale F, The Black Box Society: The Secret Algorithms That Control Money and Information (Harvard University Press 2015)

Pearl J and Mackenzie D, The Book of Why: The New Science of Cause and Effect (Allen Lane 2018)

Pease K, Repeat Victimisation: Taking Stock (Home Office 1998)

Perrow C, *Normal Accidents: Living with High Risk Technologies.* (Princeton University Press 2011)

Perry W, McInnis B, Price C, Smith S, and Hollywood J, Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations (RAND Corporation 2013).

Privacy International, 'Digital Stop and Search' (Privacy International 2018)

Privacy International, Submission to the Information Commissioner–request for an assessment notice of data brokers Experian & Equifax (PI 2018).

Press Association, ''Custody Image' Deletion Request Figures Revealed', Mail Online (12 February 2018) <http://www.dailymail.co.uk/wires/pa/article-5379353/Custody-image-deletion-request-figures-revealed.html> accessed 28 April 2019

Purshouse J and Campbell L, 'Privacy, Crime Control and Police Use of Automated Facial Recognition Technology' (2019) 3 Criminal Law Review 188

Purtova N, 'Between the GDPR and the Police Directive: Navigating through the Maze of Information Sharing in Public–Private Partnerships' (2018) 8 International Data Privacy Law 52

Rice M, 'Seizing the future: Seeking clarity of law in the search and seizure of mobile devices in Scotland', In: Proceedings of the 2019 British and Irish British and Irish Law Education and Technology Association Conference (BILETA 2019)

Robinson G, 'Stand-down and Deliver: Pre-Sentence Reports, Quality and the New Culture of Speed' (2017) 64 Probation Journal 337

Rosser G, Davies T, Bowers KJ, Johnson SD, and Cheng T, 'Predictive Crime Mapping: Arbitrary Grids or Street Networks?' (2017) 33 J Quant Criminol 569

Sales J, 'The Public Sector Equality Duty' (2011) 16 Judicial Review 1

Selbst AD and Powles J, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233.

Selbst AD, Boyd D, Friedler SA, Venkatasubramanian S, and Vertesi J, 'Fairness and Abstraction in Sociotechnical Systems' in Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19, New York, NY, USA, ACM 2019).

Singh JP, Kroener DG, Wormith JS, Desmarais SL, and Hamilton Z (eds), Handbook of Recidivism Risk/Needs Assessment Tools (Wiley Blackwell 2018)

Simon, MA, Communication de M. Alain Simon à la conférence annuelle des commissaires à la protection des données (Québec, septembre 1987), reported in Commission nationale de l'informatique et des libertés (CNIL), 8e Rapport au président de la République et au Parlement, 1987 (La Documentation Francaise 1988) ⟨https://perma.cc/2NCW-R5Q3⟩

Skitka LJ, Mosier KL, and Burdick M, 'Does Automation Bias Decision-Making?' (1999) 51 International Journal of Human-Computer Studies 991

Sommer P, Written Evidence FRS0009 to the House of Lords Science and Technology Inquiry into Forensic science and the criminal justice system (2018)

Stanton JM, 'Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors' (2001) 9(3) Journal of Statistics Education DOI: 10/gd82dx

Susskind R, *Expert Systems in Law: A Jurisprudential Inquiry* (Clarendon Press 1989)

Susskind R, *The Future of Law: Facing the Challenges of Information Technology* (Oxford University Press 1987)

The Law Society of England and Wales, Criminal justice system in crisis: Parliamentary briefing. (The Law Society 2019).

The Scottish Parliament, Justice Sub-Committee on Policing, Report on Police Scotland's proposal to introduce the use of digital device triage systems (cyber kiosks) (April 2019)

The Scottish Parliament, Justice Sub-Committee, *Official Report* (September 13 2018)

Tversky A and Kahneman D, 'Judgment under Uncertainty: Heuristics and Biases' (1974) 185 Science 1124

Urwin S, 'Algorithmic Forecasting of Offender Dangerousness for Police Custody Officers: An Assessment of Accuracy for the Durham Constabulary Model' (MSc, University of Cambridge 2016)

Ustun B, Spangher A, and Liu Y, 'Actionable Recourse in Linear Classification' in Proceedings of the ACM Conference on Fairness, Accountability and Transparency (ACM FAT*) (ACM 2018)

Veale M and Brass I, 'Administration by Algorithm? Public Management Meets Public Sector Machine Learning' in K Yeung and M Lodge (eds), Algorithmic Regulation (Oxford University Press 2019) doi:10/gfzvz8

Veale M and Edwards L, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling' (2018) 34 Computer Law & Security Review 398.

Veale M, Binns R, and Edwards L, 'Algorithms that Remember: Model Inversion Attacks and Data Protection Law' (2018) 376 Phil Trans R Soc A 20180083.

Veale M, Van Kleek M, and Binns R, 'Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making' in Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2018 (ACM Press 2018) doi:10/ct4s

Wachter S, Mittelstadt B, and Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76.

Wick MR and Thompson WB, 'Reconstructive Expert System Explanation' (1992) 54 Artificial Intelligence 33

Wiles P, Annual Report 2017:Commissioner for the Retention and Use of Biometric Material (Office of the Biometrics Commissioner 2018)

Willcocks LP and Lacity M, Service Automation Robots and the Future of Work (SB Publishing 2016)

Williams E, Norman J, and Wunsch D, 'Too Little Too Late: Assessing Vulnerability' (2009) 3 Policing 355

Yeung K, '"Hypernudge": Big Data as a Mode of Regulation by Design' (2017) 20 Information, Communication & Society 118