Royal United Services Institute
for Defence and Security Studies

BRIEFING PAPER

Data Analytics and Algorithmic Bias in Policing

Alexander Babuta and Marion Oswald

ABOUT THIS PAPER

This paper summarises the use of analytics and algorithms for policing within England and Wales, and explores different types of bias that can arise during the product lifecycle. The findings are based on in-depth consultation with police forces, civil society organisations, academics and legal experts. The primary purpose of the project is to inform the Centre for Data Ethics and Innovation's ongoing review into algorithmic bias in the policing sector. RUSI's second and final report for this project will be published in early 2020, to include specific recommendations for the final Code of Practice, and incorporating the cumulative feedback received during the consultation process.

The authors made an equal contribution to this paper, and are very grateful to Jamie Grace, Pete Fussey, Malcolm Chalmers, Cathy Haenlein and the RUSI Publications team for providing helpful feedback on an earlier draft of this report.

SUMMARY

- The use of data analytics and algorithms for policing has numerous potential benefits, but also carries significant risks, including those relating to bias. This could include unfair discrimination on the grounds of protected characteristics, real or apparent skewing of the decision-making process, or outcomes and processes which are systematically less fair to individuals within a particular group. These risks could arise at various stages in the project lifecycle.
- Algorithmic fairness cannot be understood solely as a matter of data bias, but requires careful consideration of the wider operational, organisational and legal context, as well as the overall decision-making process informed by the analytics.
- While various legal frameworks and codes of practice are relevant to the police's use of analytics, the underlying legal basis for use must be considered in parallel to the development of policy and regulation. Moreover, there remains a lack of organisational guidelines or clear processes for scrutiny, regulation and enforcement. This should be addressed as part of a new draft code of practice, which should specify clear responsibilities for policing bodies regarding scrutiny, regulation and enforcement of these new standards.

INTRODUCTION

RUSI was commissioned by the Centre for Data Ethics and Innovation (CDEI) to conduct an independent research study on the use of data analytics by police forces in England and Wales, with a focus on algorithmic bias. The primary purpose of the project is to inform CDEI's review into algorithmic bias in the policing sector.

CDEI's review will work towards a draft Code of Practice for the development, trialling and implementation of data analytics in policing, which will mitigate risk of bias and address wider legal and ethical concerns.

This briefing paper is the first of two papers to be published as part of this project. Its purpose is to outline the main issues identified by the research so far, and to offer contextual information for stakeholders providing feedback on the draft Code of Practice, due to be published by CDEI. The second RUSI paper will be published in early 2020 and will include specific recommendations for the final CDEI Code of Practice, incorporating feedback received during the consultation process.

Building on RUSI's existing work in this area, this project combines semi-structured key informant interviews, roundtables and focus groups, and a selected review of literature focused on data analytics and algorithmic bias. To date, key informant interviews have been conducted with 13 representatives of various UK law enforcement agencies, and five academics and legal experts. In addition, two roundtable events were held in London in July 2019. The first brought together 16 representatives from the commercial police technology sector and was organised in partnership with TechUK, while the second brought together 27 participants from police forces, civil society organisations, government departments, and academic and legal experts. A semi-structured interview protocol was used systematically throughout.[1] Further interviews are planned and the second paper will include cumulative insights from all interviews, discussion of the regulatory and oversight regime, and comments on the draft Code of Practice.

This briefing paper summarises the use of analytics and algorithms for policing in England and Wales, before discussing different types of bias that can arise during the product lifecycle. The purpose of this paper is not to offer solutions or recommendations as to how these risks can be addressed; this will be discussed in detail in the second paper. Instead, this briefing paper highlights the most significant gaps in the existing policy framework where future regulatory attention should be focused.

For a more general overview of the state of police technology in England and Wales, see RUSI's 2017 paper 'Big Data and Policing: An Assessment of

[1] Throughout this report, an anonymised coding system is used to refer to interview data. The prefix 'L' is used to refer to law enforcement representatives, while 'A' refers to academic and legal experts.

Law Enforcement Requirements, Expectations and Priorities'.[2] For a more detailed discussion of the types of machine learning algorithms currently in use, a useful starting point is RUSI's 2018 report 'Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges'.[3]

POLICE USE OF DATA ANALYTICS IN ENGLAND AND WALES

UK police forces collect vast amounts of digital data, but have historically lacked the technological capabilities to effectively analyse this data to improve operational effectiveness and efficiency.[4] However, police forces are increasingly adopting advanced analytical tools to derive insights from the data they collect, to inform decision-making, resource prioritisation and risk assessment in a range of contexts.

The analytical tools used by police forces increasingly employ forms of machine learning, often referred to as artificial intelligence. However, this latter description is ambiguous and poorly defined, so for the purposes of this paper the technology in question is referred to as 'machine learning'. Machine learning algorithms are currently used for various policing purposes, including: facial recognition and video analysis; mobile phone data extraction; social media intelligence analysis; predictive crime mapping; and individual risk assessment.[5] This report focuses on these latter two applications of machine learning, which are frequently referred to as forms of 'predictive policing'.[6] However, many of the same legal, ethical and policy issues apply to other uses of machine learning in policing, including those linked to classification, explanation and resource allocation.

The use of algorithms to make predictions about future crime and offending raises considerable legal and ethical questions

While the use of predictive policing tools in the UK can be traced back to at least 2004,[7] advances in machine learning have enabled the development

[2] Alexander Babuta, 'Big Data and Policing: An Assessment of Law Enforcement Requirements, Expectations and Priorities', *RUSI Occasional Papers* (September 2017).

[3] Alexander Babuta, Marion Oswald and Christine Rinik, 'Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges', *Whitehall Report*, 3-18 (September 2018).

[4] Babuta, 'Big Data and Policing'.

[5] The Law Society Commission on the Use of Algorithms in the Justice System and the Law Society of England and Wales, 'Algorithms in the Criminal Justice System', June 2019.

[6] Beth Pearsall, 'Predictive Policing: The Future of Law Enforcement', *National Institute of Justice Journal* (No. 266, 2010), pp. 16–19; Jennifer Bachner, 'Predictive Policing: Preventing Crime with Data and Analytics', IBM Center for the Business of Government, 2013; Walter L Perry et al., *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (Santa Monica, CA: RAND Corporation, 2013).

[7] Kate J Bowers et al., 'Prospective Hot-Spotting: The Future of Crime Mapping?', *British Journal of Criminology* (Vol. 44, No. 5, September 2004), pp. 641–58;

of more sophisticated systems, which are now used for a wider range of functions. The use of algorithms to make predictions about future crime and offending raises considerable legal and ethical questions, particularly concerning the risk of bias and discrimination.[8]

DOES IT WORK?

Before discussing the risks of bias arising from predictive policing technology, it is important to address the fundamental question – 'does it work?' It is beyond the scope of this paper to critically assess the (dis)advantages of a risk assessment-focused approach to resource allocation, or to discuss the semantic nuances associated with defining 'risk'. However, on the basis that police forces must target limited resources to places and people identified as posing the greatest 'risk' (of offending or victimisation), this paper's starting point is to question whether algorithmic tools are effective in assisting the police to identify and understand this risk. Effectiveness and accuracy are intrinsically linked to ethics and legality: if it cannot be demonstrated that a particular tool or method is operating effectively and with a reasonable degree of accuracy, it may not be possible to justify the use of such a tool as necessary to fulfil a particular policing function.[9]

First, in relation to predictive mapping, empirical evidence has demonstrated that the deployment of predictive mapping software could increase the likelihood of detecting future crime events when compared to non-technological methods, resulting in net reductions in overall crime rates.[10] Research shows that random foot patrolling has a negligible impact on detecting and preventing crime, because crime is not uniformly distributed in time and space.[11] By contrast, 'hotspot' policing – whereby high-risk locations are identified and patrol resources concentrated in those areas – has been shown to result in crime suppression not just at

Shane D Johnson et al., *Prospective Crime Mapping in Operational Context,* Final Report, Home Office Online Report 19/07 (London: The Stationery Office, 2007).

[8] Hannah Couchman, 'Policing by Machine', *Liberty,* January 2019.

[9] Author's telephone interview with A4, academic expert in human rights and technology, 11 July 2019.

[10] Johnson et al., *Prospective Crime Mapping in Operational Context*; Kate J Bowers et al., 'Spatial Displacement and Diffusion of Benefits Among Geographically Focused Policing Initiatives: A Meta-Analytical Review', *Journal of Experimental Criminology* (Vol. 7, No. 4, December 2011), pp. 347–74.

[11] George L Kelling et al., 'The Kansas City Preventive Patrol Experiment', Police Foundation, 1974; Lawrence W Sherman, Patrick R Gartin and Michael E Buerger, 'Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place', *Criminology* (Vol. 27, No. 1, February 1989), pp. 27–56; Shane D Johnson et al., 'Space–Time Patterns of Risk: A Cross National Assessment of Residential Burglary Victimization', *Journal of Quantitative Criminology* (Vol. 23, No. 3, 2007), pp. 201–19; Anthony A Braga, 'The Effects of Hot Spots Policing on Crime', *Annals of the American Academy of Political and Social Science* (Vol. 578, No. 1, November 2001), pp. 104–25.

the deployment location but also in surrounding areas.[12] In the UK, field trials have found predictive mapping software to be around twice as likely to predict the location of future crime as traditional intelligence-led techniques (whereby analysts manually identify future hotspots).[13]

Despite its apparent effectiveness, the use of predictive mapping software by UK police forces has been limited.[14] In many cases, its use has amounted only to short-term trials that did not result in full-scale deployment.[15]

The evidence is less clear when it comes to the accuracy of individual risk-assessment tools, largely due to a lack of research on the algorithms in use. Nevertheless, there is a large body of research dating back more than 60 years comparing the accuracy of 'unstructured' professional judgement and statistical ('actuarial') forecasting methods, which it is not possible to discuss here.[16] Various meta-analyses and systematic reviews have found that – under controlled conditions – statistical forecasting consistently outperforms unstructured professional judgement in a range of decision-making contexts, including offender risk assessment.[17]

However, experts disagree over the predictive validity of statistical risk-assessment tools.[18] Predictive validity can be understood as 'the extent to which scores on an assessment tool are able to predict some outcome

[12] Lawrence W Sherman and David Weisburd, 'General Deterrent Effects of Police Patrol in Crime "Hot Spots": A Randomized, Controlled Trial', *Justice Quarterly* (Vol. 12, No. 4, 1995), pp. 625–48; Rob T Guerette and Kate J Bowers, 'Assessing the Extent of Crime Displacement and Diffusion of Benefits: A Review of Situational Crime Prevention Evaluations', *Criminology* (Vol. 47, No. 4, November 2009), pp. 1331–68; College of Policing, 'The Effects of Hot-Spot Policing on Crime: What Works Briefing', September 2013.

[13] George Mohler et al., 'Randomized Controlled Field Trials of Predictive Policing', *Journal of the American Statistical Association* (Vol. 110, No. 512, 2015), pp. 1399–411; Kent Police Corporate Services Analysis Department, 'PredPol Operational Review [Restricted and Heavily Redacted]', <http://www.statewatch.org/docbin/uk-2014-kent-police-predpol-op-review.pdf>, accessed 14 August 2019.

[14] Conchman, 'Policing by Machine', p. 45.

[15] Conchman, 'Policing by Machine', pp. 45–61; Babuta, 'Big Data and Policing'.

[16] For further discussion, see Robyn M Dawes, David Faust and Paul E Meehl, 'Clinical Versus Actuarial Judgment', *Science* (Vol. 243, No. 4899, 1989), pp. 1668–74; William M Grove and Paul E Meehl, 'Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical–Statistical Controversy', *Psychology, Public Policy, and Law* (Vol. 2, No. 2, 1996), pp. 293–323.

[17] Paul E Meehl, *Clinical Vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (Minneapolis, MN: University of Minnesota Press, 1954); William M Grove et al., 'Clinical Versus Mechanical Prediction: A Meta-Analysis', *Psychological Assessment* (Vol. 12, No. 1, 2000), p. 19.

[18] See, for example, Stephen D Hart and David J Cooke, 'Another Look at the (Im-)Precision of Individual Risk Estimates Made Using Actuarial Risk Assessment

measure'.[19] However, if a statistical tool is used to make predictions at the individual level, the uncertainty associated with any single event probability is very large. As summarised by Alan A Sutherland and colleagues, 'predictive judgments are meaningful when applied to groups of offenders. However, at an individual level, predictions are considered by many to be imprecise'.[20] Put simply, high accuracy rates at the group level can often conceal very low accuracy rates for specific individuals or groups of individuals within that larger group. All individual predictions are associated with a confidence interval (a margin of error), which is often not taken into account when reporting the overall 'predictive accuracy' of the tool.

Academic experts interviewed for this study expressed reservations regarding the ability of algorithmic tools to predict future crime, indicated by comments such as 'there are a lot of myths around machine learning tools and what they can do. One of the things that machine learning is really terrible at is predicting rare and infrequent events, especially when you don't have loads of data'.[21] With this in mind, the more infrequent the event the tool is trying to predict, the less accurate it is likely to be. Furthermore, accuracy is often difficult to calculate, because when an individual is judged to pose a risk of offending, an intervention is typically delivered which prevents the predicted outcome from happening. Authorities cannot know what may have happened had they not intervened, and therefore there is no way to test the accuracy (or otherwise) of the prediction.

Independent, methodologically robust evaluation of trials is essential

Independent, methodologically robust evaluation of trials is essential to demonstrate the accuracy and effectiveness of a particular tool or method. If such evaluation does not demonstrate the tool's effectiveness and proportionality, continued use would raise significant legal concerns regarding whether use of the tool was justified to fulfil a particular policing function, requiring the police force to review its design and operational use. Conversely, if there is evidence that a new capability is beginning to perform well, it is important to invest in building the evidence base for its effectiveness, with processes in place for ongoing evaluation.

THE CURRENT LANDSCAPE

In England and Wales, a small number of police forces have developed machine learning algorithms to assess reoffending risk for known offenders

Instruments', *Behavioral Sciences and the Law* (Vol. 31, No. 1, January/February 2013), pp. 81–102.

[19] Mia Debidin (ed.), *A Compendium of Research and Analysis on the Offender Assessment System (OASys) 2006–2009* (London: Ministry of Justice, 2009), p. 78.

[20] Alan A Sutherland et al., 'Sexual Violence Risk Assessment: An Investigation of the Interrater Reliability of Professional Judgments Made Using the Risk for Sexual Violence Protocol', *International Journal of Forensic Mental Health* (Vol. 11, No. 2, 2012), p. 120.

[21] Academic expert during Interdisciplinary Roundtable event hosted by RUSI and the CDEI, London, 25 July 2019.

in the force area, to inform prioritisation of operational activity and to assist decision-making at the entry point to the criminal justice system. For instance, Durham Constabulary's Harm Assessment Risk Tool uses random forest forecasting (a form of supervised machine learning) to classify individuals in terms of their likelihood of committing a violent or nonviolent offence over the next two years.[22] The purpose is to assist officers in assessing offenders' eligibility to participate in the Checkpoint Programme, a voluntary out-of-court disposal scheme designed to reduce reoffending by addressing the underlying factors causing individuals to engage in crime.[23] Avon and Somerset Constabulary uses similar technology to assess factors such as likelihood of reoffending, likelihood of victimisation/vulnerability, and likelihood of committing a range of specific offences. Through an app on their mobile devices, neighbourhood officers can instantly access the risk profiles for each offender registered in the force area, which are recalculated on a daily basis.[24] West Midlands Police are developing a similar offender assessment system as part of their Data Driven Insights project,[25] while Hampshire Constabulary is developing a machine learning predictive tool to assess risk of domestic violence offending.[26]

The current technological landscape was described by one police officer interviewed as a 'patchwork quilt, uncoordinated and delivered to different standards in different settings and for different outcomes'.[27] However, the use of analytics and algorithms by police forces in England and Wales is likely to grow in both scale and sophistication in the coming years. It is essential to build a stronger evidence base on the effectiveness and reliability of different systems, and to develop a clearer legal, policy and regulatory framework to ensure proportionate and ethical use of this increasingly powerful technology.

[22] Sheena Urwin, 'Algorithmic Forecasting of Offender Dangerousness for Police Custody Officers: An Assessment of Accuracy for the Durham Constabulary Model', unpublished thesis, University of Cambridge, 2016.

[23] Durham Constabulary, 'Checkpoint', <https://www.durham.police.uk/Information-and-advice/Pages/Checkpoint.aspx>, accessed 14 August 2019.

[24] Lina Dencik et al., 'Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services', Cardiff University, 2018.

[25] See West Midlands Police and Crime Commissioner and West Midlands Police Ethics Committee, papers from committee meeting, April 2019, <https://www.westmidlands-pcc.gov.uk/archive/april-2019/>, accessed 14 August 2019.

[26] Petros Terzis, Marion Oswald and Christine Rinik, 'Shaping the State of Machine Learning Algorithms Within Policing', workshop report, University of Winchester, June 2019.

[27] Authors' telephone interview with L6, representative of UK law enforcement agency, 10 July 2019.

RISK OF BIAS IN ALGORITHMIC POLICE DECISION-MAKING

This overview focuses on the application of the law of England and Wales to algorithmic tools in policing, together with key ethical, practical and policy issues. It is beyond the scope of this paper to assess the underlying legal basis for the use of algorithmic tools for policing, including positive obligations under Article 2 of the European Convention on Human Rights (ECHR)[28] (right to life), or the potential need for primary legislation, although these are crucial issues that should be considered in parallel.[29] In this project, 'bias' is understood not only as indicating incomplete or unrepresentative datasets, but instead takes a broader view of the whole decision-making process. This project approaches 'bias' as it relates to:

- Outcomes or processes which are systematically less favourable to individuals within a particular group where there is no justification for such difference, thereby creating new targeted groups not necessarily linked to protected characteristics.[30]
- Direct or indirect discrimination on the grounds of protected characteristics, in contravention of the Equality Act 2010[31] and Article 14 of ECHR.
- Real or apparent skewing of the decision-making process such that it is or would appear unfair, including where the decision had in practice been predetermined.[32]

There is a risk that focusing on 'fixing' bias as a matter of 'data' may distract from wider questions of whether a predictive algorithmic system should be used at all in a particular policing context.[33] The context of austerity raises questions about the justifiability of using algorithmic tools to increase efficiency, where they may not be necessary if more resources were available.[34] Therefore, the risk of bias is considered not only in terms of data inputs and outputs, but from a contextual perspective, within the

[28] 'Convention for the Protection of Human Rights and Fundamental Freedoms', 1950.

[29] For an assessment of the legal basis of live facial recognition, see Pete Fussey and Daragh Murray, 'Independent Report on the London Metropolitan's Police Service's Trial of Live Facial Recognition Technology', The Human Rights, Big Data and Technology Project, July 2019.

[30] Nicol Turner Lee, Paul Resnick and Genle Barton, 'Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms', *Brookings,* 22 May 2019.

[31] 'Equality Act 2010 (UK)', c.15.

[32] This reflects, in particular, administrative law natural justice (rules of fair administrative procedure applicable to public bodies) and Article 6(1) of the ECHR.

[33] Julia Powles, 'The Seductive Diversion of "Solving" Bias in Artificial Intelligence', *Medium*, 7 December 2018.

[34] Author's telephone interview with L1, representative of UK law enforcement agency, 1 July 2019.

main stages of a project lifecycle. As one senior police officer commented, 'The big issue in policing is not the technology. It's what the military call the "capability stack" – the combination of the technology, the people and the processes that need to be considered'.[35]

Several legal frameworks and codes of practice are relevant to the development and deployment of predictive algorithms in England and Wales, including:

- Data protection, specifically Part 3 of the Data Protection Act 2018.[36]
- Prohibited discrimination under the Equality Act 2010 and the public sector equality duty.[37]
- Obligations pursuant to the ECHR and section 6 of the Human Rights Act 1998.[38]
- Statutory responsibilities regarding coercive and investigatory powers.[39]
- Requirements pursuant to the Criminal Procedure and Investigations Act 1996 relating to investigation, prosecution and disclosure of evidence.[40]
- The duties of the police within the common law.
- Administrative law principles applicable to lawful public sector decision-making.[41]
- College of Policing Authorised Professional Practice,[42] including the Code of Ethics[43] and Management of Police Information guidelines.[44]

[35] Authors' telephone interview with L8, representative of UK law enforcement agency, 23 July 2019.

[36] 'Data Protection Act 2018 (UK)', Part 3; Council of the European Union, 'Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016', *Official Journal of the European Union* (L 119).

[37] 'Equality Act 2010 (UK)', c.15, s149.

[38] 'Human Rights Act 1998 (UK)'; 'Convention for the Protection of Human Rights and Fundamental Freedoms'.

[39] For instance, within the Police and Criminal Evidence Act 1984, Regulation of Investigatory Powers Act 2000, Protection of Freedoms Act 2012, and Investigatory Powers Act 2016.

[40] 'Criminal Procedure and Investigations Act 1996 (UK)', c.25.

[41] Marion Oswald, 'Algorithmic-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power', *Philosophical Transactions of the Royal Society A* (Vol. 376, No. 2128, 2018); Jamie Grace, '"Algorithmic Impropriety" in UK Policing?', *Journal of Information Rights, Policy and Practice* (Vol. 3, No. 1, 2019); Jennifer Cobbe, 'Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making', *Legal Studies* (forthcoming).

[42] College of Policing, 'Authorised Professional Practice', <https://www.app.college.police.uk/app-content/>, accessed 23 August 2019.

[43] College of Policing, 'Code of Ethics', <https://www.college.police.uk/What-we-do/Ethics/Ethics-home/Pages/Code-of-Ethics.aspx>, accessed 23 August 2019.

[44] College of Policing, 'Authorised Professional Practice: Management of Police Information', <https://www.app.college.police.uk/app-content/information-

The project of which this paper is part assesses the risk of bias in algorithmic decision-making in relation to these frameworks.

TYPES OF POTENTIAL BIAS IN THE PROJECT LIFECYCLE

Problem and Solution Identification

An increased emphasis on the preventive and public safety aspects of the police's role, coupled with a significant reduction in resources since 2010, has led to a perceived need to prioritise based on data-driven assessments of risk.[45] However, bias can arise when choosing which specific crime problems will be the subject of such data-driven assessments. At the 'problem identification' phase, predictive technological solutions have been criticised for focusing on low-level 'nuisance' crime, or on areas with high crime levels and thus poor neighbourhoods.[46] One academic expert suggested that bias may arise via the police's conception of a problem (and thus input data would reflect only this conception), for example the conception of a 'gang' framed around a single demographic.[47]

Despite general cuts in police funding since 2010, specific funds for 'digital transformation' have been made available, such as the Police Transformation Fund, 'creating strong incentives for forces to frame the development around digital technology to receive further central support'.[48] This may create a bias in favour of digital solutions, and the risk of modelling complex social issues in an overly simplistic way,[49] without equal consideration of non-technological measures. This raises questions regarding necessity and proportionality and data minimisation issues if the use of digital solutions results in increased data collection to build, test and monitor a model. An academic expert suggested that, where adoption of predictive policing occurs rapidly, 'what is often missing … is a testing and assessment of the risks and long-term benefits that these systems may provide'.[50] Another police respondent pointed to 'a risk that we procure capabilities based on opinion-based decisions rather than evidence-based decisions'.[51] With this

management/management-of-police-information/>, accessed 23 August 2019.

[45] Dencik et al., 'Data Scores as Governance'.

[46] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York, NY: Penguin, 2016).

[47] Author's telephone interview with A5, academic expert in policing policy and practice, 11 July 2019.

[48] The Law Society Commission on the Use of Algorithms in the Justice System and The Law Society of England and Wales, 'Algorithms in the Criminal Justice System', p. 13.

[49] Author's telephone interview with A4, academic expert in human rights and technology, 11 July 2019.

[50] Author's interview with A3, academic expert in information law and data protection, London, 28 June 2019.

[51] Author's telephone interview with L1, representative of UK law enforcement agency, 1 July 2019.

in mind, the choice of whether to implement a particular new technological capability may itself be subject to bias.

Design and Testing

Previous research has demonstrated how choices about defining target variables (the thing that the model is trying to predict) and input variables (the factors taken into account during computation) may adversely impact protected classes.[52]

Algorithms that are trained on police data 'may replicate (and in some cases amplify) the existing biases inherent in the dataset',[53] such as over- or under-policing of certain communities, or data that reflects flawed or illegal practices[54] (raising further issues regarding the requirement of accuracy pursuant to the Data Protection Act).[55] A police officer interviewed commented that 'young black men are more likely to be stop and searched than young white men, and that's purely down to human bias. That human bias is then introduced into the datasets, and bias is then generated in the outcomes of the application of those datasets'.[56] The effects of a biased sample could be amplified by algorithmic predictions via a feedback loop,[57] whereby future *policing* is predicted, not future crime.[58] Another officer commented that 'we pile loads of resources into a certain area and it becomes a self-fulfilling prophecy, purely because there's more policing going into that area, not necessarily because of discrimination on the part of officers'.[59]

In addition to these biases inherent in police data, individuals from disadvantaged sociodemographic backgrounds are likely to engage with public services more frequently, meaning the police often have access to more data relating to these individuals, which may in turn lead to them being calculated as posing a greater risk. Seemingly relevant data could also serve as proxies for protected characteristics (and sensitive data under the Data Protection Act) – for example, home address or number of stop-and-searches

[52] Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact', *California Law Review* (Vol. 104, 2016), p. 671.

[53] Babuta, Oswald and Rinik, 'Machine Learning Algorithms and Police Decision-Making'.

[54] Rashida Richardson et al., 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice', *New York University Law Review Online* (forthcoming).

[55] 'Data Protection Act 2018 (UK)', c. 12, ss. 37, 38.

[56] Author's telephone interview with L1, representative of UK law enforcement agency, 1 July 2019.

[57] Frederik Zuiderveen Borgesius, 'Discrimination, Artificial Intelligence and Algorithmic Decision-Making', Council of Europe, 2018.

[58] Danielle Ensign et al., 'Runaway Feedback Loops in Predictive Policing', *Proceedings of Machine Learning Research* (Vol. 81, 2018), pp. 1–12.

[59] Author's telephone interview with L2 and L3, representatives of UK law enforcement agency, 4 July 2019.

as proxies for ethnicity. In terms of the use of protected characteristics themselves, discrimination claims could be brought by individuals scored 'negatively' in comparison to others of different ages or gender, and race discrimination claims brought where postcode is a proxy for race,[60] although the claimant would need to show less favourable treatment. This may prove complicated when 'less favourable treatment' could amount to *not* being subject to a particular intervention, for instance an out-of-court disposal scheme such as Durham Constabulary's Checkpoint Programme.

Complications arise when a model must account for legitimate differences in offending across demographics. For instance, men commit crime at significantly higher rates than women and are more likely to commit violent offences.[61] Age is also strongly correlated with offending: it is well known that offending tends to peak in the teenage years and then decay over time.[62]

Scores of 'predictive accuracy' derived from retrospective validation are representative only of that sample, and do not account for model shrinkage when the algorithm is applied to new, unfamiliar data.[63] While typically presented as an individual-level prediction, the output from an offender assessment tool can be better understood as a group-level classification, describing the extent to which group members conform to a certain 'profile' identified in historic data, raising questions of fairness under data protection and human rights principles.[64] Percentage accuracy rates may also disguise high percentage false positives or negatives which could negatively impact

[60] Robin Allen and Dee Masters, 'Algorithms, Apps & Artificial Intelligence: The Next Frontier in Discrimination Law', updated for the Public Law Project session entitled 'AI Justice: Artificial Intelligence Decision-Making and the Law' on 16 October 2018, Cloisters, 2018.

[61] See Jennifer Schwartz et al., 'Trends in the Gender Gap in Violence: Reevaluating NCVS and Other Evidence', *Criminology* (Vol. 47, No. 2, June 2009), pp. 401–25.

[62] David P Farrington, 'Age and Crime', *Crime and Justice* (Vol. 7, 1986), pp. 189–250. For further discussion, see, for example, Stephen D Hart, Christine Michie and David J Cooke, 'Precision of Actuarial Risk Assessment Instruments: Evaluating the "Margins of Error" of Group v. Individual Predictions of Violence', *British Journal of Psychiatry* (Vol. 190, No. S49, 2007), s63; Nancy R Cook and Nina P Paynter, 'Performance of Reclassification Statistics in Comparing Risk Prediction Models', *Biometrical Journal* (Vol. 53, No. 2, March 2011), p. 237.

[63] Chris Webster, Quazi Haque and Stephen J Hucker, *Violence Risk: Assessment and Management: Advances Through Structured Professional Judgement and Sequential Redirections*, 2nd Edition (Chichester: John Wiley & Sons, 2013), p. xiii; Kevin S Douglas, Melissa Yeomans and Douglas P Boer, 'Comparative Validity Analysis of Multiple Measures of Violence Risk in a Sample of Criminal Offenders', *Criminal Justice and Behavior* (Vol. 32, No. 5, October 2005), pp. 501–02.

[64] Laurens Naudts, 'How Machine Learning Generates Unfair Inequalities and How Data Protection Instruments May Help in Mitigating Them', in Ronald Leenes et al. (eds), *Data Protection and Privacy: The Internet of Bodies* (Hart, 2018).

on certain groups.[65] Furthermore, limiting the testing phase to statistical accuracy could mean that risks related to automation bias are missed, if the testing phase only assesses the algorithm itself rather than the overall decision-making process the tool is feeding into.[66]

The CDEI-commissioned Landscape Summary on bias in algorithmic decision-making identified proprietary (closed-source) software as a factor likely to impede scrutiny of bias.[67] A number of academic experts interviewed pointed to potential issues with off-the-shelf machine learning tools where there is no access to the training data (requiring beta testing with new data to assess bias).[68] It was noted that a system's procurer requires considerable scope to inspect and challenge a tool, almost undermining the point of outsourcing.[69] Furthermore, outsourcing risks a loss of control over the provenance of input datasets, which may have been developed for non-law enforcement purposes, in different law enforcement environments or according to different standards.[70] As one police officer commented: 'We wouldn't outsource our arrests, we wouldn't outsource our intelligence functions. Why are we outsourcing analysts?'[71]

Deployment

Deployment of a predictive algorithm may result in important contextual information being disregarded, which could introduce systematic bias into the decision-making chain in an effort to 'streamline' the process. For instance, a decision on detention post-arrest based on a predictive output could be skewed if the algorithm only uses data confirming risk as opposed to data demonstrating the opposite.[72]

[65] Julia Angwin et al., 'Machine Bias', *ProPublica,* 23 May 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, accessed 14 August 2019.

[66] Alex Albright, 'If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions', The Little Dataset, 15 July 2019, <https://thelittledataset.com/2019/07/15/if-you-give-a-judge-a-risk-score/>, accessed 14 August 2019.

[67] Michael Rovatsos, Brent Mittelstadt and Ansgar Koene, 'Landscape Summary: Bias in Algorithmic Decision-Making', CDEI, 2019.

[68] Author's telephone interview with A1, academic expert in data protection, privacy and surveillance policy, 28 June 2019.

[69] Author's interview with A2, academic expert in computer science and data protection, London, 28 June 2019.

[70] See Orla Lynskey, 'Criminal Justice Profiling and EU Data Protection Law: Precarious Protection from Predictive Policing', *International Journal of Law in Context* (Vol. 15, Special Issue 2, June 2019), pp. 162–76.

[71] Author's telephone interview with L4, representative of UK law enforcement agency, 9 July 2019.

[72] Dencik et al., 'Data Scores as Governance'.

Bias may also occur in the way that the human decision-maker adheres to or deviates from the algorithm's prediction or insight.[73] A factor which may arise during deployment (thus necessitating attention during design) is the risk of automation bias, the tendency to over-rely on automated outputs and discount other correct and relevant information.[74] Inappropriate fettering of discretion through use of an algorithm may make the resulting decision unfair,[75] or de facto an automated one.[76] As explained by one police officer interviewed: 'Officers often disagree with the algorithm. I'd expect and welcome that challenge. The point where you don't get that challenge, that's when people are putting that professional judgement aside'.[77] However, another officer noted that 'professional judgement might just be another word for bias', explaining that 'whenever we have to decide an outcome there's always an opportunity for bias'.[78] It is essential that the correct balance is struck to ensure due regard is paid to the insights derived from analytics, without making officers over-reliant on the tool and causing them to disregard other relevant factors. Adequate training focused on cognitive bias and fair decision-making would appear essential to ensure officers are able to consistently achieve the correct balance.

Furthermore, while the risk of cognitive bias is often used to argue in favour of the statistical approach, a 'risk score' is potentially highly prejudicial to the decision-maker.[79] One police respondent referred to the output getting 'into the head' of the officer.[80] Quantifications of risk are also liable to misinterpretation by those not directly involved in the assessment.[81] For example, a 'low-risk' label could be interpreted to mean that an individual requires no further monitoring or intervention. Yet such 'low-risk' individuals may have specific needs that should be addressed as part of a bespoke risk-management plan. Such individuals may then fail to receive the support to prevent them returning to problematic behaviour.

[73] Albright, 'If You Give a Judge a Risk Score'; Bo Cowgill, 'The Impact of Algorithms on Judicial Discretion: Evidence from Regression Discontinuities', working paper, 2018, <http://www.columbia.edu/~bc2656/workingpapers.html>, accessed 14 August 2019.

[74] Danielle Keats Citron, 'Technological Due Process', *Washington University Law Review* (Vol. 85, No. 6, 2008).

[75] Oswald, 'Algorithmic-Assisted Decision-Making in the Public Sector'.

[76] 'Data Protection Act 2018 (UK)', Part 3.

[77] Author's telephone interview with L4, representative of UK law enforcement agency, 9 July 2019.

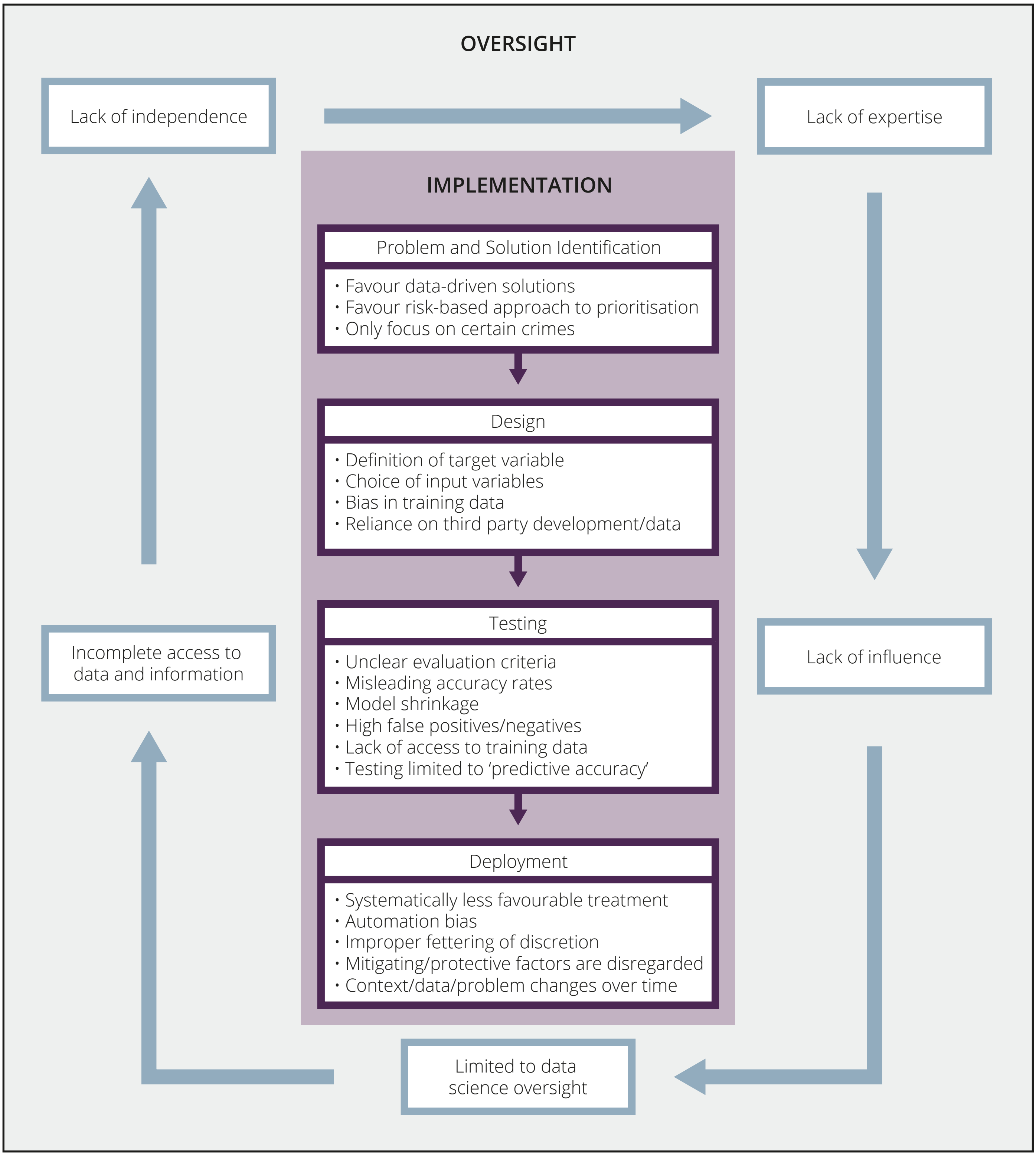
[78] Author's telephone interview with L1, representative of UK law enforcement agency, 1 July 2019.

[79] David J Cooke, 'More Prejudicial than Probative', *Journal of the Law Society of Scotland* (Vol. 55, No. 1, 2010), pp. 20–23; David J Cooke and Christine Michie, 'Violence Risk Assessment: From Prediction to Understanding – Or From What? To Why?', in Caroline Logan and Lorraine Johnstone (eds), *Managing Clinical Risk: A Guide to Effective Practice* (Abingdon: Routledge, 2013), p. 10.

[80] Authors' telephone interview with L6, representative of UK law enforcement agency, 10 July 2019.

[81] Cooke, 'More Prejudicial than Probative'.

Figure 1: Risk of Bias in Implementation and Oversight of Police Algorithms


OVERSIGHT
Lack of independence
Lack of expertise
Lack of influence
Limited to data science oversight
Incomplete access to data and information
IMPLEMENTATION
Problem and Solution Identification
• Favour data-driven solutions
• Favour risk-based approach to prioritisation
• Only focus on certain crimes
Design
• Definition of target variable
• Choice of input variables
• Bias in training data
• Reliance on third party development/data
Testing
• Unclear evaluation criteria
• Misleading accuracy rates
• Model shrinkage
• High false positives/negatives
• Lack of access to training data
• Testing limited to 'predictive accuracy'
Deployment
• Systematically less favourable treatment
• Automation bias
• Improper fettering of discretion
• Mitigating/protective factors are disregarded
• Context/data/problem changes over time

Source: The authors.

EMERGING FINDINGS

Interviews conducted to date evidence a desire for clearer national guidance and leadership in the area of data analytics, and widespread recognition and appreciation of the need for legality, consistency, scientific validity and oversight. It is also apparent that systematic investigation of claimed benefits and drawbacks is required before moving ahead with full-scale deployment of new technology.[82] As one law enforcement practitioner commented, 'there's as much value in understanding what doesn't work, as what does',[83] but to achieve this, controlled space for experimentation is required, recognising that 'policing is about dealing with complexity, ambiguity and inconsistency'.[84]

Lessons can be learned from recent trials of live facial recognition, particularly concerning the need to demonstrate an explicit legal basis for the use of new technology, the need for clearer guidance relating to trials and evaluation, and the importance of meaningful public engagement during the development and testing phase. The development of a draft Code of Practice provides an opportunity, not only to consider bias, but to improve understanding of the application of data analytics in different contexts, and of methods of assessing potential benefits and intrusions. It will be incumbent on users to evidence such assessments when determining whether use of a particular tool can be deemed 'necessary', in order to decide whether there are less intrusive means of achieving the same policing aim.[85]

Any new code of practice for algorithmic tools in policing should establish a standard process for model design, development, trialling, and deployment, along with ongoing monitoring and evaluation. It should provide clear operationally relevant guidelines and complement existing authorised professional practice and other guidance in a tech-agnostic way.[86] Existing surveillance codes and related inspections were suggested by a number of interviewees as a potential model. The new code should ensure sufficient attention is paid to meeting legal and ethical requirements throughout all stages of the product lifecycle, from project inception through to model procurement, development and testing, including ongoing tracking and mitigation of discrimination risk when the tool is deployed operationally,

[82] See Albert Meijer and Martijn Wessels, 'Predictive Policing: Review of Benefits and Drawbacks', *International Journal of Public Administration* (Vol. 42, No. 12, 2019), pp. 1–9.

[83] Author's telephone interview with L7, representative of UK law enforcement agency, 18 July 2019.

[84] Authors' telephone interview with L8, representative of UK law enforcement agency, 23 July 2019.

[85] Author's telephone interview with A4, academic expert in human rights and technology, 11 July 2019.

[86] Author's telephone interview with L1, representative of UK law enforcement agency, 1 July 2019.

and oversight of the ultimate decision-making process the analytical insights are feeding into.[87]

A new code should specify clear roles and responsibilities regarding scrutiny, regulation and enforcement, including the roles of the College of Policing, the National Police Chiefs' Council, Her Majesty's Inspectorate of Constabulary and Fire & Rescue Services and the Home Office, and potentially other regulatory bodies such as the Information Commissioner's Office and Investigatory Powers Commissioners. The code should also establish standard processes for independent ethical review and oversight to ensure transparency and accountability and facilitate meaningful public engagement before tools are deployed operationally.

ABOUT THE AUTHORS

Alexander Babuta is a Research Fellow in National Security Studies at RUSI. He leads the Institute's research on policing, intelligence and surveillance, with a focus on the use of emerging technologies for security purposes.

Marion Oswald is the Vice-Chancellor's Senior Fellow in Law at the University of Northumbria, an Associate Fellow of RUSI and a solicitor (non-practising). She is Chair of the West Midlands Police and Crime Commissioner and West Midlands Police Ethics Committee, a member of the National Statistician's Data Ethics Advisory Committee and an executive member of the British and Irish Law, Education and Technology Association.

[87] Authors' telephone interview with L6, representative of UK law enforcement agency, 10 July 2019.

About RUSI

The Royal United Services Institute (RUSI) is the world's oldest and the UK's leading defence and security think tank. Its mission is to inform, influence and enhance public debate on a safer and more stable world. RUSI is a research-led institute, producing independent, practical and innovative analysis to address today's complex challenges.

Since its foundation in 1831, RUSI has relied on its members to support its activities. Together with revenue from research, publications and conferences, RUSI has sustained its political independence for 188 years.

The views expressed in this publication are those of the authors, and do not necessarily reflect the views of RUSI or any other institution.

Published in 2019 by the Royal United Services Institute for Defence and Security Studies. RUSI is a registered charity (No. 210639).

Royal United Services Institute
for Defence and Security Studies
Whitehall
London SW1A 2ET
United Kingdom
+44 (0)20 7747 2600
www.rusi.org

RUSI is a registered charity (No. 210639)