Shaping the state of machine learning algorithms within policing

A report to discuss the outcomes of a consultation workshop that took place on Thursday 6 June 2019 at the Royal United Services Institute, Whitehall, London with the purpose of moving the debate on **machine learning algorithms** within **policing** from high level principles towards a detailed **discussion** and **critique** of existing knowledge and research.

Run by the Centre for Information Rights and supported by the University of Winchester through the Higher Education Innovation Fund.

Authors Petros Terzis, Marion Oswald and Christine Rinik, University of Winchester
Date 26 June 2019

UNIVERSITY OF WINCHESTER
CENTRE FOR INFORMATION RIGHTS

Contents

Introduction and Acknowledgements

The workshop, run by the Centre for Information Rights, University of Winchester and the Royal United Services Institute (RUSI), followed the 2018 report on machine learning and policing. Authored by the same institutions, this report recommended a multidisciplinary approach to address the 'real-world' challenges raised by the use of machine learning algorithms for police decision-making. The workshop was organised by the Centre for Information Rights, University of Winchester in collaboration with RUSI, and was supported by the Higher Education Innovation Fund.

The agenda brought together participants from law enforcement, government, policy-makers, the legal sector and academia to share and debate issues relating to design and deployment of machine learning within operational law enforcement environments, notions of fairness within computer science, and proposals for regulation and oversight.

The remarks are those of the report authors and do not necessarily reflect the views of the workshop participants.

Executive Summary and Takeaways

- In the pursuit of deploying machine learning, the technology's complexity might lure us away from identifying the core characteristics and the roots of the **problem** that we are trying to solve.
- Full **transparency** should be regarded as a prerequisite for a meaningful discussion of the use of machine learning algorithms within public administration and especially law enforcement and criminal justice.
- A **Data Protection Impact Assessment and Equality Impact Assessment** should be drafted and adopted at a very early stage, during the scoping of the project.
- The concept '**Fairness**' is subject to different disciplinary interpretations. More work is needed to bring those interpretations together and to ensure that work done to uphold fairness is grounded in an understanding of underlying legal principles as well as particular context.
- Algorithmic **explainability** and or/transparency are not sufficient conditions to guarantee the fairness of a model and such a procedure cannot be transformed into a box-ticking course of action.
- The quest for algorithmic regulation might lead us towards 'reinventing the wheel' whilst there seems to be an adequate arsenal of legal principles already. The problem is not the absence of law but its **multi-diversity** and lack of **specificity**.
- The role of the **annual spending review** should be re-examined, as emphasis on yearly technology spending may have a detrimental effect on proper planning and implementation processes.
- There is a need for a new **Code of Practice** (taking into consideration existing codes such as those linked with the Police and Criminal Evidence Act, voluntary codes such as those linked with stop-and-search, and proposed methods such as 'Algocare') linked with the use of algorithmic decision making systems and an appropriate independent authority to oversee their use.

Feedback from delegates

At the end of the workshop, participants were asked to complete a questionnaire determining whether the presentations and discussions were likely to have an **impact** on their way of thinking, planning and working. According to the submissions we received, the event influenced thinking in a number of areas related to:

1. the necessity for an earlier engagement -during the tasking process- with legal and ethical questions and demands;
2. the development of sectoral and governmental codes of practice, independent oversight frameworks and guidelines;
3. the source, relevance and validity of data and the risk of 'dirty data';
4. the implications for policy and practice of differing definitions of fairness.

The structure

The workshop was divided into three sessions. Each session started with speakers' remarks and was followed by discussants' comments as well as roundtable conversations where participants debated suggested questions. At the end of every session, one participant from each of the six roundtables presented the outcomes of the group reflections.

The questions addressed in each of the three session were as follows:

Session 1: Using machine learning in risk assessment and prediction in operational environments: practicalities of tool justification, design and deployment

- What factors should be considered when setting objectives and specifications at the start of a machine learning tool development process?
- What does an 'effective' machine learning tool look like in a policing context?
- What are the main challenges of introducing a machine learning tool into an existing decision-making process?

Session 2: Notions of fairness in machine learning design

- Is everyone talking about different things when we discuss 'fairness' and 'bias'?

Session 3: Law, regulation and oversight and Lessons from the US

- Is the current oversight regime within policing fit to govern an algorithm-assisted future?

Session One:

Using machine learning in risk assessment and prediction in operational environments: practicalities of tool justification, design and deployment

Opening Talk *by Chief Supt David Powell, Hampshire Constabulary & Bogdan Mecu, Deloitte*

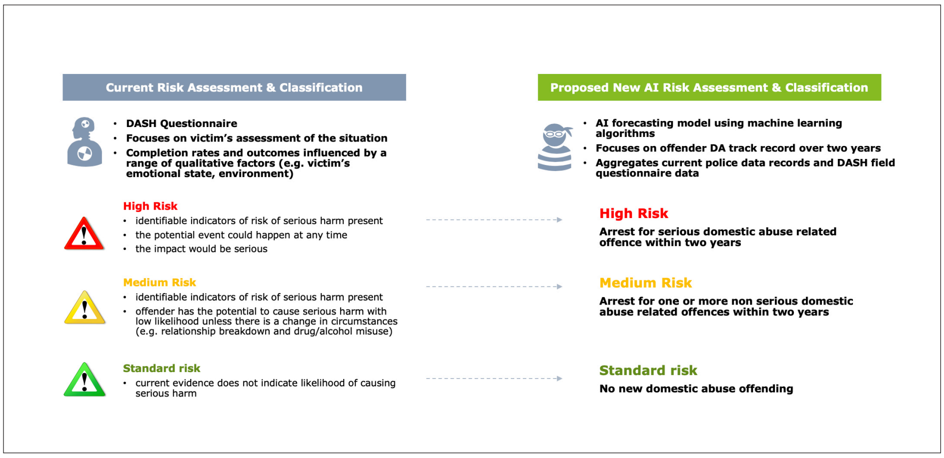
Domestic Abuse (DA) is one of the few areas in policing where the response is driven almost exclusively by a formal risk assessment tool. The Domestic Abuse, Stalking and Honour Violence Risk, Identification, Assessment and Management model (DASH) was implemented in 2009. From 2014 onwards, Hampshire Constabulary has registered a **200% increase** in domestic abuse occurrences. There are several reasons for this namely; improved recording of crime, increased focus and reporting, and a change in the definition of domestic abuse to include a wider cohort of potential victims. Whilst DASH has some well recognised strengths there are a number of inherent weaknesses that undermine its efficacy as a risk assessment tool[1]:

- It is not supported by an independent peer reviewed evidence base;
- It claims to be non-predictive but the assessment of risk drives a response based on probability i.e. high risk will mean police and partner agencies will allocate resources and operational focus to safeguard a victim on the basis that it is more likely than not that an event of serious harm will occur. On this basis the risk assessment tool is indeed a predictor.
- It assesses risk of serious harm only and not general re victimisation;
- It is not time bound;
- It generates a disproportionate amount of false positives particularly at high risk;
- It has low reliability (different versions of the truth);
- Environmental factors impact field deployment.

---

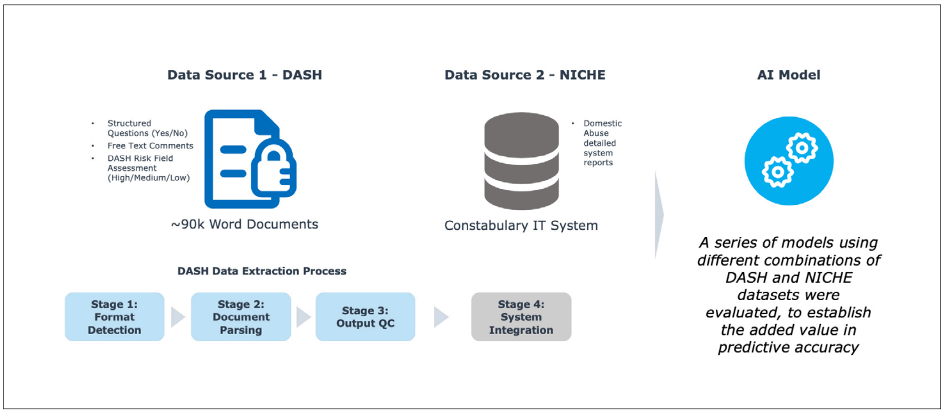
1 Emily Turner, Juanjo Medina, Gavin Brown, Dashing Hopes? The Predictive Accuracy of Domestic Abuse Risk Assessment by Police, The British Journal of Criminology, azy074, https://doi.org/10.1093/bjc/azy074

To improve the assessment of risk to help better safeguard victims, research has focused on developing an advanced Artificial Intelligence (AI) model to forecast the future dangerousness of domestic abuse perpetrators. The forecasting model does not seek to replace DASH but to improve it by providing an additional perpetrator based risk classification. The DASH risk assessment represents the voice of the victim in that it is a victim led assessment of the risk posed to them. In developing the AI tool a key focus was trying to include the voice of the victim in the model and ensuring that any DA risk assessment in the field continued to be focused on listening and recording victims' views. The table below illustrates the existing risk assessment process and the proposed AI model.



Picture taken from speakers' slides presentation

From a technical point of view, the new AI model is based on two datasets: extracted data from DASH risk assessments and additional existing police records (NICHE):



Picture taken from speakers' slides presentation

The text from these datasets is translated into numeric format and the AI forecasting model learns to classify the risk of domestic violence re-offending, from multiple data points. Example data points are shown below although these are not the full list:

- Offender (age, gender, arrested, suspected, self-harm/suicide warnings);
- Victim (age, gender);

- Incident (serious, violence, sexual, offender count, victim count, first domestic occurrence, top 30 postcode -district level);
- Criminal record (Intimate Occurrences, Murder Offences, Intimate Serious Offence, Intimate Violence Offence); and
- DASH risk assessment (Incident Rating, Prior DASH High Ratings, 28 Yes/No Questions, Risk Assessment).

Currently the new AI model is at the Proof of Concept only stage. The time allocated for the presentation did not allow for a deeper examination and explanation of the initial results, given the breadth of the subject matter, and additional and continuous evaluation will continue using a variety of methods including as regards false positives and false negatives.

There is currently cautious support for using this new tool in the field – as a 'friend' to officers' discretion[2] giving them another dataset to consider in conjunction with their own initial field assessment based on DASH and their professional knowledge and experience. Building on this acceptance and the potentially promising primary results, the next step forward is to develop a minimum viable version of the new AI tool to test in practice and give further insights into the potential efficacy of such an approach to risk classification in domestic abuse. A decision to support a minimum viable trial has not yet been approved and is dependent on a number of further considerations. The project team are focused on proceeding carefully with development given the complexity of implementation issues that lie ahead. Below is a non-hierarchical list of further issues to be addressed;

- Implementation decision & stakeholder support;
- Governance framework;
- Ownership & accountability (risk & intellectual property);
- IT Integration & process design;
- User centric approach to deployment (e.g. confidence, training;)
- Legal & ethical considerations;
- On-going model maintenance (e.g. re-training, feedback loop);
- Outcome evaluation;
- National scale & lessons learned.

2 Presentation of David Powell based on the presenter's research findings.

Discussion

The participants expressed their general appreciation of the openness around discussing this model in detail as full transparency should be regarded as a prerequisite for a meaningful discussion on the use of machine learning algorithms within public administration and especially law enforcement and criminal justice. Furthermore, taking into account purpose limitation, the presented system seems to benefit from the fact that it was set up for single purpose. Whilst accepting the relatively high level nature of the presentation and the necessity for a more detailed description of the model, it appears that it may be striking the right balance without overloading the system with unnecessary data.

The discussion included reflection on the information relevant to decision-making in domestic abuse cases, including the voice of the victim (which the model aimed to reflect through the DASH information), and how the model could support the professional decision-making of the officer via existing processes. It was noted that it was important that consideration was given to potential consequences of translating text-based information into numeric representations readable by a machine learning model. The risk of 'automation bias' was highlighted as a key consideration for any future implementation processes and procedures. The importance of training and empowering user confidence in appropriate use of the tool was discussed, as was the need for a clear governance regime. It was noted that data protection and equality impact assessments, as well as other evaluation models, can be crucial in highlighting data processing, tool design, implementation and evaluation issues that should be addressed.

It was also noted that the annual funding for police technology projects may not support the need for multi-year development and considered testing and trialling of proposed tools. The deployment of machine learning tools and the processing of data by such tools potentially engages Article 8 of the European Convention on Human Rights (ECHR) and therefore consideration should be given to how the principles of necessity and proportionality will be made out, including by comparison with the status quo.

The following points were put forward in answer to the discussion questions:

1. **What factors should be considered when setting objectives and specifications at the start of a machine learning tool development process?**

a) What **problem** are we trying to tackle? What makes machine learning a suitable tool for this purpose?

b) What are the lawful constraints? What guidance is there in statute (including requirements **to undertake a Data Protection Impact Assessment (DPIA) very early**/at the start of scoping a project) and from a regulator such as the Information Commissioner's Office (ICO)?

c) What are the **user requirements**? **Organisational requirements** too?

d) **Where is the money coming from**? Proprietary concerns - private sector funded development? Development funded by 'transformation funding' from government sources?

2. What does an 'effective' machine learning tool look like in a policing context?

a) It would accurately identify risks of re-offending (but query how 'accurate' should be defined);

b) It would be used in a way that is conscious of the 'fragmentation of knowledge' problem;

c) University-police partnerships and relationships might need to be consciously governed by university-model ethical processes;

d) Clear measurements of an effective contribution;

e) The smoothest adoption/blending of the use of these tools in police roles.

3. What are the main challenges of introducing a machine learning tool into an existing decision-making process?

a) Funding;

b) Lack of training;

c) Changing ways of working;

d) Gaps in information;

e) Over-reliance or over-optimism;

f) Ignorance of evidence base (e.g. AI-informed policing of domestic violence might not reduce domestic violence as much as AI-informed public health interventions);

g) Ensuring that legal and ethical factors are considered at design stage.

Questions for further research:

- How might the design of DASH or of any other pre-existing type of model undermine the development of a new/supplementary systems?
- How can we best estimate the utility of machine learning tools?

Session Two:

Notions of fairness in machine learning design

Opening talk *by Reuben Binns, University of Oxford*

The 'fairness' discussion and the challenge for 'de-biased' algorithms still dominate the interdisciplinary debate around the use of algorithms in the public sphere. The primary origin of the problem is the different meanings that each discipline endows 'fairness' and 'bias' with. Despite these disparities in the 'essence' of these principles, there seems to be an agreement at least as to the causes of the genesis of bias.

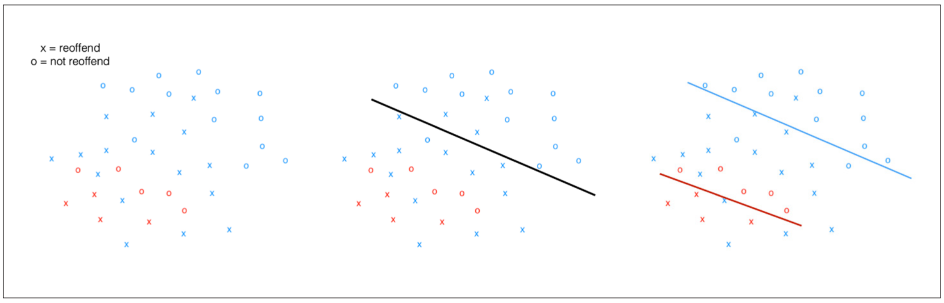
The first source of bias is the high or low percentage of a minority's representation within a dataset. In such a case, building different models for different sub-sets might alleviate the problem. Secondly, there is ample evidence demonstrating that human biases are encoded within the historic data that are used to train the algorithms. In such a context, the complexity of machine learning structures can lead the model to inadvertently infer protected characteristics. For example, today 'black people are now nine times more likely to be stopped and searched for drugs despite using drugs at a lower rate than white people'[3] . Thirdly, unequal base rates is another source of bias not necessarily due to existing social bias but the question of whether it should be included in the model is something that needs special attention.

Fairness for computer scientists is a 'different beast' and can be evaluated in three different stages: pre-processing where the data are 'massaged' to eliminate biases, in-processing where we intervene with the learning logic of algorithms at the training stage and post-processing where we change the trained model altogether. The measures for such an evaluation are divided between two categories: group measures and individuals measures.

3 Michael Shiner, Zoe Carre, Rebekah Delsol and Niamh Eastwood, The Colour of Injustice: 'Race', drugs and law enforcement in England and Wales, available at: https://www.release.org.uk/sites/default/files/pdf/publications/The%20Colour%20of%20Injustice.pdf

Group measures

Anti-classification



Sides from speaker's presentation referencing the cited authors and the **Stanford Computational Policy Lab Tutorial**

The model is fair if it does not use **protected characteristics** or **proxies** from which protected characteristics can be inferred. For example a model that uses Qualifications, Work Experience, Attendance, Gender, Race, Education Institution, Postcode as data is not 'fair' as it uses protected characteristics (Gender and Race) as well as potential proxies (Education Institution and Postcode).

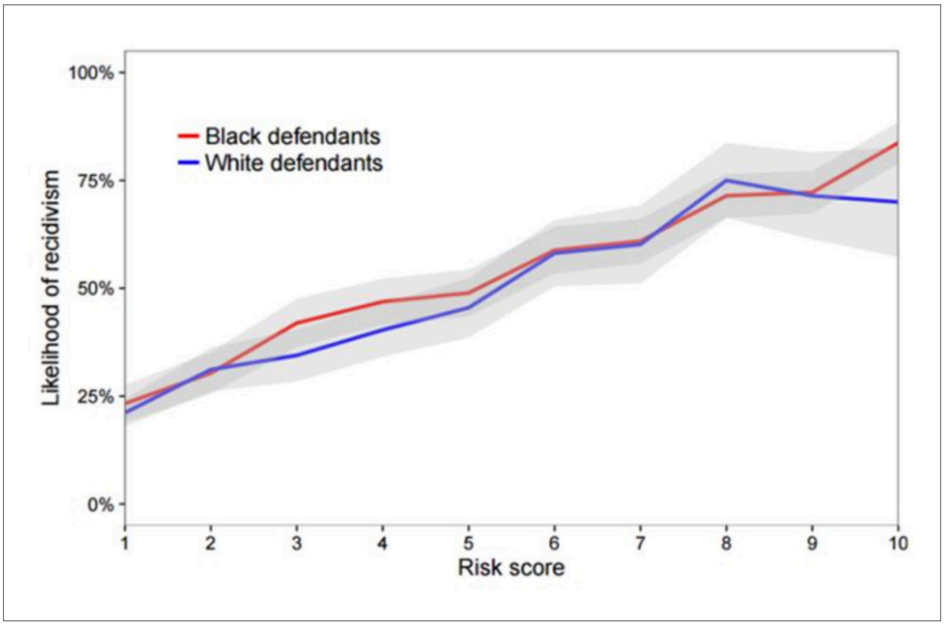
Classification outcome/error parity

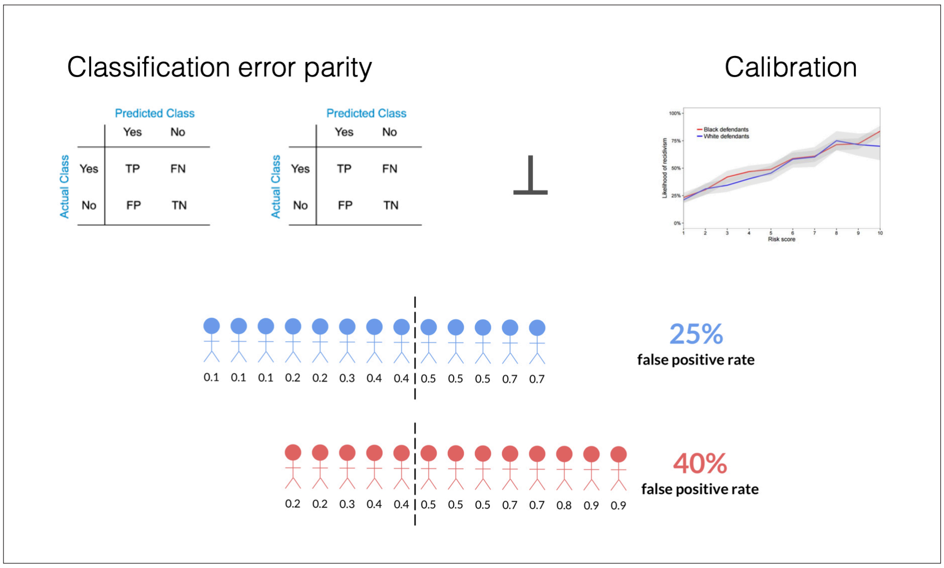
The model is fair if protected groups receive equal proportion of positive outcomes[4] or equal proportion of errors. For example in job application tools, where 40% of male applicants get offers, 40% of female applicants should get offers while in a predictive model false positives should occur equally between men and women.

Calibration

For any specific risk score band, split people who are given that score by protected groups, should see equal calibration. For example, of those given a particular risk score S, S% should result in the predicted outcome. Calibration should be equal between protected groups.

4 Calders, T. & Verwer, S. Data Min Knowl Disc (2010) 21: 277. https://doi.org/10.1007/s10618-010-0190-x



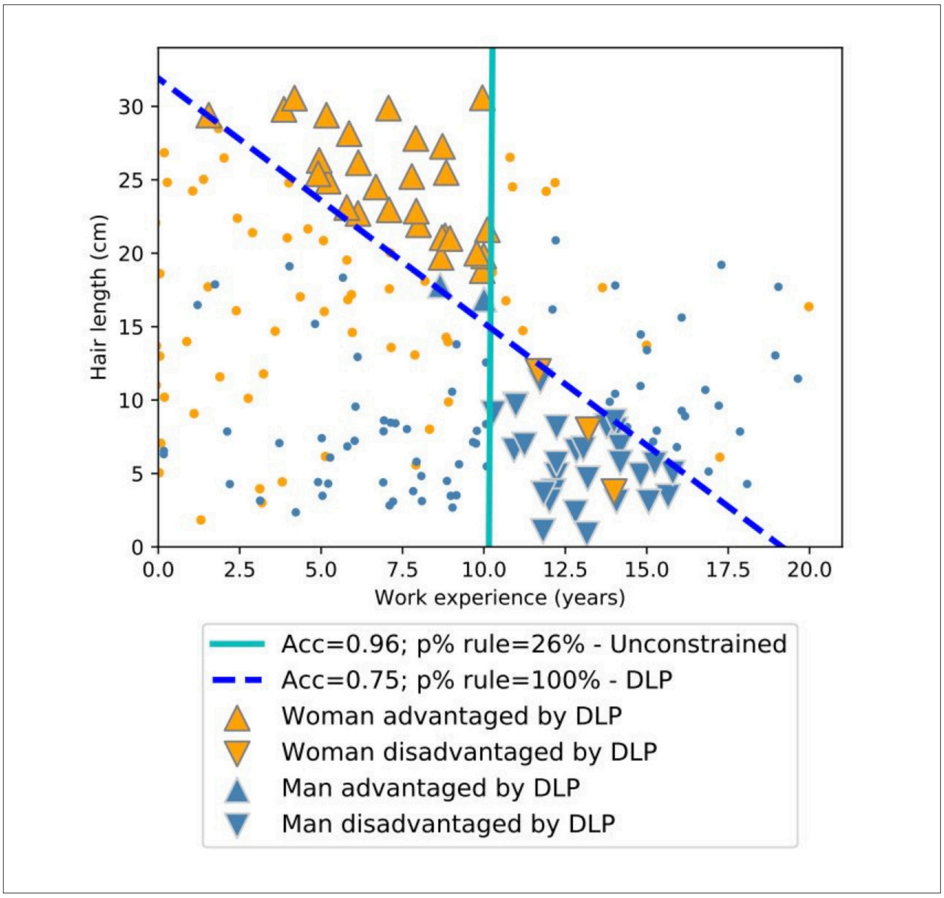


Slides from speaker's presentation referencing the cited authors and the **Stanford Computational Policy Lab Tutorial**

One of the biggest problems with group measures is their mutual **incompatibility**. For example, classification error parity is incompatible with calibration[5]. Equally correcting models to satisfy classification parity **and** anti-classification leads to perverse outcomes. Hence, all fairness definitions based on infra-marginal statistics (e.g. FPR, FNR) will suffer from this problem[6]. As risk distributions between populations differ, infra-marginal statistics will also differ.

5 Alexandra Chouldechova, *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments*, (2017) , arXiv:1703.00056

6 Corbett-Davies, Sam & Goel, Sharad. (2018), *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*, arXiv:1808.00023



Slides from speaker's presentation referencing the cited authors and the **Stanford Computational Policy Lab Tutorial**

Individual fairness

'Individual fairness' on the other hand signifies the features that are relevant for the specific task. Thus, pairs of individuals who are similar in task-relevant features should be treated the same. Such a process is not without problems. Firstly, it can potentially get quite complicated when there are complex combinations of characteristics. Secondly, the very denotation of 'task-relevant' can be tricky as it touches upon policy-making issues and hence there needs to be a task-relativity metric.

However there are lines that can be drawn to distinguish between 'accepted' and 'non-accepted' variables. Such a line ought to go through the ability of the individual to change the attributes he/she is held responsible for. Judging people based on attributes upon which they can have no impact is inherently questionable. *Equality of Opportunity* already referred to in '*Fairness through awareness*'[7] and more recently in '*On the Long-term Impact of Algorithmic Decision Policies: Effort Unfairness and Feature Segregation through Social Learning*'[8] is used as a regime to separate acceptable grounds for discrimination (e.g. relevant

7 Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, Rich Zemel, *Fairness through awareness* (2012), arXiv:1104.3913

8 Hoda Heidari & Vedant Nanda, *On the Long-term Impact of Algorithmic Decision Policies: Effort Unfairness and Feature Segregation through Social Learning (2019)* arXiv:1903.01209

personal qualities) from unacceptable grounds (e.g. irrelevant qualities and circumstances for which the individual should not be held morally accountable). It further dictates that for any context you are dealing with you will first need to define your **circumstances** and the **effort** and to ensure that the outcome is only a function of the effort.

Equality of Opportunity appears to resolve tensions between individual and group fairness measures as similar individuals are those who share effort-based similarity. Hence if the effort variables are influenced by circumstance, then similar individuals will share similar quantile across circumstance types. At the same time, Equality of Opportunity stumbles upon certain group metrics. From a statistical point of view error parity and calibration reflect different assumptions about what people are accountable for. While the latter assumes that people with same predicted label are equally accountable for that label, for the former people with same true label are equally accountable for it. However, these assumptions tend to neglect that some people become disadvantaged by choosing to prevent disadvantage to others (e.g. caregivers) and disadvantages arising from such choices maybe be unjust even if the choice was 'free'. These are choices inherently influenced and constrained by social context and institutional structures. Thus, albeit free, such choices still generate the question whether they are genuine and autonomous enough to be used as a criterion to ground a judgment on an individual.

Discussion

Police data is partial. It has problematic quality and it is by no means self-evident that it is possible to guarantee fairness for there will always be trade-offs. In this context, we need to give as much attention to how things are used, as to how they are prepared. And although it is accepted that explainability is a *conditio sine qua non* for fairness, it is not *per se* enough for algorithms to be considered 'fair'. Apart from the algorithms, **fairness itself also needs explanation**.

Serious concerns were raised regarding the data literacy as a problem within all levels in policing. In an era of immense information there is a constant need for critical thinking to assess information effectively and put algorithms in the right place at the right time. Besides, relying on automated-decision making systems generates the risk of automation bias which results in certain information being neglected and might end up rendering police officers less skilled[9]. At the same time, the human-in-the-loop of law enforcement might find himself/herself in a difficult position when dealing with the outcome of a machine due to the reasonable fear of doing the wrong thing or the temptation of over-thinking and doing nothing in the end.

---

9 Lisanne Bainbridge, Ironies of automation, Automatica, Volume 19, Issue 6, 1983, Pages 775-779, ISSN 0005-1098, https://doi.org/10.1016/0005-1098(83)90046-8.

In the context of law enforcement, fairness is often interpreted as procedural fairness. However, there is a lack of public awareness regarding procedural 'unfairness' generally. Moreover, when it comes to policing context, 'bias' is not perceived as a *stricto sensu* disparity in the data but can be a systematic issue around the organisation, and equally, discrimination may result from intervention impact or priorities. Priority setting, however, is never done just by the police itself. On the contrary, it is infused through political and institutional input as a result of policy trade-offs and targets.

Assessment of fairness could depend upon intervention e.g. even if only 1% is error that 1% might have significant effect on individual if intervention is proved to be detrimental. The reason why such an assessment is that complicated is because we cannot agree on methods to approach this principle and there are no agreed best practices. Simultaneously, society and academia seem to hold algorithms to higher standards of fairness or freedom from bias than it does for human beings.

The following points were put forward in answer to the discussion questions:

Is everyone talking about different things when we discuss 'fairness' and 'bias'?

Yes- Fairness can be interpreted as justice; as non-bias; as equality; as transparency etc. Statistical definitions of fairness and procedural standards of fairness will never be written in stone the same way because the context-dependent nature of concepts of fairness will always metamorphose subject to and driven by more context.

Questions for further research:

- What if the choices that comprise the effort towards an outcome are driven by unfair social circumstances?
- In the context of policing, who is making decisions around trade-offs in the model?
- If the model only assists in allocating benefits, not detriments, is unfairness alleviated?
- How can we avoid the 'human crumple zone'[10] bearing responsibility but not agency?
- If data is built on human ground truth and judgment, and that judgment atrophies, where does new data come from?

---

10 Elish, M. C., *Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction* (pre-print) (March 1, 2019). Engaging Science, Technology, and Society (pre-print). Available at SSRN: https://ssrn.com/abstract=2757236 or http://dx.doi.org/10.2139/ssrn.2757236

Session Three:

A: Law, regulation and oversight

Opening Talk *by Christine Rinik, University of Winchester and Jamie Grace, Sheffield Hallam University*

In the context of policing, machine learning algorithms are automated mathematical formulas which operate based on a series of instructions to create an output. By using training data, they develop pattern recognition whose output can be utilised as a 'predictor' for future behaviour. This prediction is based on correlation identified within the specific dataset and does not suggest causation in the way this principle has been interpreted within the policing and criminal justice framework. Thus, these algorithms are meant to support (not replace) human decision-making.

Apart from the bias issue that has already been mentioned, there are additional challenges surrounding the use of algorithms in this context that require consideration prior to implementation. Some of these issues include the following:

Accuracy

- How is accuracy measured and using what criteria?
- How do we deal with the difficulty of measuring accuracy if the prediction results in intervention which potentially prevents the forecasted behaviour occurring?

Transparency

- Is the 'black box' inevitable and if so is algorithmic decision-making compatible with the ECHR Art 6 -- **right to a fair trial**?
- **Auditability** – is it possible or necessary to be able to retroactively recreate how the output was calculated?
- Complete **source-code disclosure** – is this necessary or desirable?
- **Intelligibility**– can the user explain how the model operates in practice?

Impact on professional judgment

- Could the use of algorithmic decision-making systems unintentionally encourage **abdication of responsibility by the human user of the tool**?
- Could they undermine **professional judgment**?

- How can we best ensure individual officers are enabled and assisted by an algorithmic tool?
- What form should the output of a machine learning tool have in order to be most helpful to the user?

On the other hand, there are potential benefits to reap from the use of such systems. Firstly, they provide a proactive and time-saving method to identify those at highest risk for re-offending. Secondly, it has been argued that the same systems that can reinforce human bias can are also capable of minimising such bias whilst also subsequently improving the accuracy of the decisions. All the above render algorithms a potential 'supportive friend to the discretion' of the officer[11]. Several recommendations have been made in this regard:

- Develop codes of practice governing police trials of machine learning tools;
- Educate and empower users to understand, interpret and properly weigh the output from machine learning tools;
- Establish minimum standards of technical transparency;
- Make available the list of variables included in the model;
- Establish interdisciplinary working group(s) to review/develop machine learning requirements within policing context—police, computer science, law, ethics;
- Retain public sector control of software and data sets;
- Ensure constant attention and vigilance to review the outputs of the machine learning algorithms[12].

Regulating algorithms will not be a simple matter. However before embarking on a research and policy endeavour to shape an adequate regulatory framework, it would be helpful to consider whether there is indeed a lack of law in the field. With current domestic law illuminated by the law of the European Union and the European Court of Human Rights is there really an absence of relevant law? Or is it more a problem of application of that law to new uses of technology? Inevitably, there lies the role that Brexit will play. The more the UK stays within a community whose norms are constantly evolving and to whom courts are aligned, the more our arsenal of principles regarding the regulation of algorithms will be solidified.

A regulatory compass for police departments that has been built in literature as a self-regulation tool and was subsequently favourably cited and acknowledged

11 Presentation of David Powell based on the presenter's research findings.

12 Alexander Babuta, Marion Oswald and Christine Rinik (2018) *Algorithms, Predictive Policing and Criminal Justice Decision-Making*, Royal United Services Institute Whitehall Report 3-18 https://rusi.org/publication/whitehall-reports/machine-learning-algorithms-and-police-decision-making-legal-ethical.

by institutions and organizations is the ALGOCARE guidance[13]. Developed in the aftermath of an evaluation undertaken on the 'Harm Assessment Risk Tool' (HART), currently used by Durham Constabulary, ALGOCARE sets out key questions to be answered in relation to any proposed deployment of algorithmic decision-making tools within law enforcement. In November 2018, a version of 'ALGOCARE' was listed for adoption and promotion to police forces in England and Wales by the National Police Chiefs' Council.

Overall, the benefits to regulating algorithmically-informed decision-making in a balanced way are clear in the long term. Great technological leaps forward can also have transformational constitutional effect for the communal good, as Hildebrandt has explained[14]:

'While the printing press first allowed the rule by law (the sovereign using written codes as a means to rule his subjects), it later enabled the rule of law (the internal division of sovereignty that separates the enactment of legal rules by the legislator from their interpretation in a court of law).'

B: Lessons from the US

Opening Talk *by Rashida Richardson, AI Now Institute*

In the US there are currently no national or state standards on how police data is recorded. Police are gatekeepers to information that are needed to understand their practices and there are few political and institutional incentives for self-monitoring, auditing, or reform. Police data reflects police practices and policies, not actual crime rates. The Los Angeles Police Department (LAPD), for example, misrecorded 14,000 serious assaults as minor offenses from 2005 to 2012. This error was not discovered until 2015, by which time LAPD had already begun its work with the predictive policing company PredPol, though there is no evidence to confirm whether this erroneous data was used in the system. Hence, the origin of the problem is the dirty data. The term is broadly used to encompass among others[15]:

---

13 Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes (2018) *Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality*, Information & Communications Technology Law, 27:2, 223-250.

14 Mireille Hildebrandt, 'Profiling and AmI', in Kai Rannenberg, Denis Royer and Andre Deuker (eds.), *The Future of Identity in the Information Society: Challenges and Opportunities*, 2009, Springer, p.300.

15 Richardson, Rashida and Schultz, Jason and Crawford, Kate, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice* (February 13, 2019). New York University Law Review Online, Forthcoming. Available at SSRN: https://ssrn.com/abstract=3333423

- data that is derived from or influenced by corrupt, biased, and unlawful practices including data that has been intentionally manipulated or 'juked';
- data that is distorted by individual and societal biases;
- data generated from the arrest of innocent people who had evidence planted on them or were otherwise falsely accused;
- data from calls for service or incident reports that reflect false claims of criminal activity;
- data from intentionally distorted police records, such as the systemic manipulation of crime statistics to try to promote particular public relations, funding, or political outcomes.

In such a context, data encompasses and reflects information that is neither apolitical nor ahistorical. Thus, inevitably one might ask why are predictive/AI tools targeted at particular types of crime and not, for example, at white collar crime. Either way, it is persuasively argued that reforms that do not require restrictions or prohibitions of dirty data will not address the problem. There is great need for an empowered and independent authority to assess and address dirty data in policing. The New York City Automated Decision Systems Task Force is expected to publish a report on a range of issues relating to government use of automated decision systems in December 2019.

Discussion

As the debate grows the degree of interdisciplinarity should increase and lessons from governance studies should flow into the debate, to join the legal and ethical discussion. Journalism, scholarship, and activism lead the way with a largely developing momentum and widely emerging synergies. However, we need to have a solid understanding of the particularities of each cooperating field. For example, when we refer to the notions of governance and accountability, we ought to always bear in mind the hierarchical structure of the police.

There are undeniable concerns around stigma, improper profiling, and punitiveness that need to be better explored and hopefully resolved. However the quest for a solution to these problems should not obscure our broader vision. Indeed, the question of the need for political ideology to potentially re-focus machine learning tools toward detecting or predicting white-collar crime, portrays a broader societal concern: the call for a different prioritization regarding the use of machine learning tools.

In an age of economic austerity in the UK and growing inequalities, predictive or explanatory analytics could serve as a powerful tool for advocating structural economic change in favour of increased investment into the root causes of crime.

Today prisons often fail to prevent reoffending and often result in poor mental health outcomes, drug addiction, insufficient support for people with learning needs and violence. Moreover, the youth are often let down due to cuts to youth services, drug treatment and school budgets, and pressures on underfunded children's services and mental health support. Shaping the ethical and legal direction of predictive analytics could transform society's perspective on criminal justice – towards a compassionate set of public health solutions and away from divisive retributive policies. The police will be a vital partner in any such agenda, but questions should be asked about who is best to lead such a reform. In order to understand developing algorithmic practices within the police, there is the need for a central audit of systems in development or operation.

The following points were put forward in answer to the discussion questions:

Is the current oversight regime within policing fit to govern an algorithm-assisted future?

There is a need for a new Code of Practice (taking into consideration existing codes such as those linked with the Police and Criminal Evidence Act, voluntary codes such as those linked with stop-and-search, and proposed methods such as 'ALGOCARE') associated with the use of algorithmic decision making systems and an appropriate independent authority to oversee their use.

Questions for further research:

- The role of the annual spending review should be reviewed, as emphasis on yearly technology spending may have a detrimental effect on proper planning and implementation processes over the longer term.
- The preventive and risk management activities of the police can overlap with other agencies. The use of algorithmic tools in connection with this agenda suggests the need for a wider debate about the role of the police and other agencies with crime prevention roles.

UNIVERSITY OF
WINCHESTER
CENTRE FOR INFORMATION RIGHTS

Sparkford Road, Winchester,
Hampshire SO22 4NR

Tel: +44 (0) 1962 841515
enquiries@winchester.ac.uk