EJLT European Journal of **Law and Technology**

A three-pillar approach to achieving trustworthy and accountable use of AI and emerging technology in policing in England and Wales:

Lessons from the West Midlands data ethics model

Marion Oswald*

Abstract

As the first of its kind in UK policing, the West Midlands Police and Crime Commissioner and West Midlands Police data ethics committee is an ongoing experiment in scrutinising and advising upon AI policing projects proposed for real operational environments, with the aim of putting people's rights at the heart of technological development. Using a qualitative action research approach akin to an 'observing participant', this paper suggests that lessons can be learned from the committee's activities in three main areas: i) the contribution to effective accountability in respect of ongoing data analytics projects; ii) the importance of the legal and scientific aspects of the interdisciplinary analysis; and iii) the role of necessity and the human rights framework in guiding the committee's ethical discussion.

The big themes underpinning the committee proceedings demonstrate the operationalisation of many of the key factors that must be considered in the human rights necessity test. The technical and statistical aspects of policing AI cannot, and should not, be isolated from the legal, contextual, operational and ethical considerations, as each will influence the other, and thus how technology is evaluated. It is important however that laws applicable to specific policing activities are not overlooked, such as those relating to stop-and-search and its application to algorithmic tools.

A three-pillar approach could contribute to achieving trustworthy and accountable use of emerging technologies in UK policing: first, governing law plus guidance and policy interpreted for the relevant context; secondly, standards, both ethical standards attached

---

* Dr Marion Oswald, Northumbria University, Newcastle. Considerable thanks go to the two anonymous peer-reviewers for their insightful and constructive comments and advice.

to personal responsibility and scientific standards; and thirdly, people at every level within policing who are committed to accountability; all of which should be subject to rolling independent oversight. In order to ensure a stable three-pillar approach however, there needs to be fewer generalisations and more specifics as regards the application of relevant law to the deployment of emerging technologies, recognising the importance of a human rights-based approach, combined with agreed scientific standards to help us decide whether things 'work' in the policing context.

The West Midlands committee has provided additional and transparent positive pressure on the force to improve – productive challenge - as well as a certain level of validation for the projects that have progressed. A national model based on the West Midlands prototype could contribute to necessary assurance and monitoring and the development of specific policy, provided that the status, resourcing and operational challenges identified in this paper are addressed, and subject to linkage with appropriate regulation and enforcement.

Keywords: Policing, technology, ethics, human rights.

Introduction

The policing technology landscape in England and Wales has been described as 'very confused' with 'no single regulatory voice' and much uncertainty as to whether the police have 'sufficient public trust to take things forward on their own without being accused of marking their own homework'.[1] England and Wales alone has forty-three Home Office police forces with differing approaches to technological developments, and a range of regulators, oversight and other public bodies with an interest in this area. Securing an effective regime to oblige police forces to account for their actions, decisions and omissions, and thus to demonstrate trustworthiness, in respect of the development and use of technology therefore represents a considerable challenge. Rowe argues that '[i]t might be possible for democratic oversight mechanisms to develop to ensure that the tools and methods of big data policing are subject to public scrutiny, but…they would have to be significantly transformed since governance, ethics and accountability largely continue to be oriented to policing practices that developed in the C19th.'[2]

Against this background, the data ethics committee established by the West Midlands Police and Crime Commissioner (PCC: the political oversight of the police in England and Wales) and West Midlands Police (WMP) has been operating for three years (referred to

---

[1] Police practitioner interview with author, 2020.
[2] Michael Rowe, *Policing the Police: Challenges of Democracy and Accountability* (Policy Press, 2020), 73.

as 'the committee').[3] The committee's terms of reference were written by Tom McNeil, then Strategic Adviser to the PCC, following an extensive consultation exercise and assessment of multiple ethics bodies' terms. This specialist independent body sits alongside WMP's Data Analytics Lab (the Lab), providing transparent advice to the PCC and Chief Constable on both Lab projects and national policing technology initiatives.

Projects scrutinised by the committee fall within the rather generic and over-used term of artificial intelligence (AI), in this context including advanced data analytics, predictive policing, natural language processing, network analysis and facial recognition. Policing tools for 'intelligent' analysis of disparate digital data sources often link to a preventative, rather than reactive, approach to harm[4] and the emerging concept of a 'public health approach' to policing (categorised by prevention, population-level interventions, partnership working and the interpretation and sharing of data),[5] and an interest in addressing significant problems such as disproportionality and intelligence gaps. Deploying technology as a simple answer to complex problems, however, risks 'the kind of rapid unravelling of community trust, support, cooperation and civil liberties that neighbourhoods have witnessed in the past.'[6] As van Brakel comments, algorithmic tools 'can be both about care and control'; the focus may be upon 'cost-efficiency and management' or upon building community trust and diversion of individuals from crime and prison.[7] Algorithmic tools can furthermore 'contribute to defining policing practices and the role of police officers, what they can and cannot do.'[8]

As the first of its kind in UK policing, the WMP committee is an ongoing experiment in scrutinising and advising upon AI policing projects proposed for real operational environments, with the aim of putting 'people's rights…at the heart of the Lab's work.'[9] I suggest therefore that lessons can be learned in three main areas:

> i) the contribution to effective accountability in respect of the Lab's undertakings;
>
> ii) the importance of the legal and scientific aspects of the interdisciplinary analysis; and

---

[3] https://www.westmidlands-pcc.gov.uk/ethics-committee/. As at the date of writing, the author chairs the committee. This paper is written in a personal capacity and not as a representative of the West Midlands PCC, West Midlands police or fellow members of the committee.

[4] Alexander Babuta and Marion Oswald (2020) 'Data Analytics and Algorithms in Policing in England and Wales: Towards A New Policy Framework' *RUSI Occasional Paper*.

[5] Helen Christmas and Justin Srivastava 'A public health approaches in policing: a discussion paper' College of Policing (2019) https://paas-s3-broker-prod-lon-6453d964-1d1a-432a-9260-5e0ba7d2fc51.s3.eu-west-2.amazonaws.com/s3fs-public/2021-02/public-health-approaches.pdf.

[6] John L.M. McDaniel, Ken G. Pease 'Introduction' in *Predictive Policing and Artificial Intelligence*, eds John L.M. McDaniel, Ken G. Pease (Routledge 2021) 23.

[7] Rosamunde van Brakel 'Rethinking predictive policing: Towards a holistic framework of democratic algorithmic surveillance' in *The Algorithmic Society,* eds Marc Schuilenburg, Rik Peeters (Routledge 2021) 106, 112.

[8] n7, 110.

[9] https://www.westmidlands-pcc.gov.uk/ethics-committee/.

iii)   the role of necessity and the human rights framework in guiding the committee's ethical discussion.

This paper therefore aims to make a key contribution to the literature on accountability and policing technology through its analysis of the committee's unique operations. It builds upon Yeung et al.'s call for more applied research to develop techniques and systems by which human rights norms are considered during technological design, development and implementation, in ways that provide '*genuinely* ethical AI'.[10] The three-pillar framework proposed by this paper – law, standards, and accountable, skilled and empowered people – has emerged through a growing understanding of the policing operational environment, and the benefits and drawbacks of the committee scrutiny process.

The paper is structured as follows. After commenting on methodological challenges, I outline the 'big themes' that emerge from 18 months of committee proceedings, which demonstrate that the technical and statistical aspects of policing AI cannot, and should not, be isolated from the legal, contextual, operational and ethical considerations, as each will influence the other, and thus how technology is evaluated. Its work has also drawn attention to a number of challenges and opportunities relating to the status, function, structure and practicalities of an advisory or scrutiny body which are summarised and discussed (Section 1).

I then comment upon the committee's interdisciplinary analysis and argue that while ethics has its place, it is not all we need for accountability and trustworthiness (Section 2). Policing technology innovation and the required investment of public funds must be based solidly on the boundaries of the law and good science. Although not a substitute, ethical consideration and debate can help to establish the moral underpinning for the use of new technologies, bringing to the surface underlying inequalities or difficult choices between particular aims or values, and ensuring that ethical issues that are not captured by a 'legal' human rights perspective are considered. The big themes demonstrate, however, that the committee is in practice operationalising many of the key factors that must be considered in the human rights necessity test, and could be said to be exercising an 'experimental proportionality' approach.[11] (Section 3) It is important however that laws applicable to specific policing activities are not overlooked, and I discuss as an example the legislation and case-law relating to stop-and-search and their application to algorithmic tools (Section 4), followed by analysis of the importance of scientific validity to the question of necessity and proportionality and the use of police powers (Section 5).

Finally, I draw the lessons learned together by proposing a **three-pillar approach** to achieving trustworthy and accountable use of AI in policing in England and Wales which I

---

[10] Karen Yeung, Andrew Howes and Ganna Pogrebna 'AI Governance by Human Rights-Centered Design, Deliberation, and Oversight: An End of Ethics Washing' in *The Oxford Handbook of Ethics of AI*, eds. Markus D. Dubber, Frank Pasquale and Sunit Das (Oxford 2020).

[11] Marion Oswald and Jamie Grace, 'The COVID-19 contact tracing app in England and "experimental proportionality"' *Public Law* (January 2021) 27-37.

suggest is generalisable to UK policing and potentially to other public sector contexts and jurisdictions:

- application of **law** aided by guidance and policy for the relevant context;
- **standards**, both scientific standards and ethical standards attached to personal responsibility; and
- **people**, who are skilled and experienced, responsible and accountable;

all of which should be subject to rolling independent scrutiny, and regulation/enforcement, for which the West Midlands model could provide a prototype (Section 6).

Methods

The development of this paper could be defined as a form of action research. I am participating in the committee activity as at date of writing (taking action) and in this paper, researching the action and its consequences by way of critical reflection, which may feed back into further implementation and refinement, thus contributing to addressing any theory-practice gap.[12] However, as chair of the committee, I could be criticised for being too close to the process and therefore likely to be overly supportive of it, a co-producer of outputs rather than a disinterested observer. Although my role is independent, I am involved in agenda-setting, recruitment of new members, long-term planning and promoting the practical impact of the committee's activities. Kislov highlights four dilemmas that 'qualitative action researchers' have to manage:

- 'Conflicting identity: being 'too academic' for practitioners and 'not academic enough' for fellow researchers;
- Compromising research rigour in order to quickly produce results fed back to non–academic partners;
- Achieving a balancing act between being critical (in an academic way) and constructive (in a practice–oriented way);
- Maintaining your own voice while truthfully reflecting the (often conflicting) voices of multiple stakeholders.'[13]

I recognise these dilemmas, in particular the issue of conflicting identity, a significant and enduring issue for researchers who dedicate time to applying knowledge and expertise to improve policy objectives and who may therefore be misunderstood or even mistrusted by fellow academics. As Kislov comments, the overarching challenge is how to maintain 'embeddedness' – in the sense of achieving enough understanding of the context and the workings of the police to enable change – alongside 'critical distance'.

Kaminski described a research role of 'observing participant' (in contrast to a participant observer or a participant role), defined by two key factors: the observing participant enters

---

[12] O'Leary's cycles of action research: O'Leary, Z. (2004). *The Essential Guide to Doing Research*. London: Sage.

[13] Roman Kislov, 'Going native in order to make a difference? Tensions of longitudinal participatory research' (2019) *BMJ Open.* vol. 9 (Suppl 1), O14. https://doi.org/10.1136/bmjopen-2019-QHRN.14.

the process through a similar social process to its other members and is subject to similar rules; and conducts research as if he or she was a researcher. 'An ideal OP lives through his/her social role, impassively registers randomly generated personal experience, and applies available data gathering techniques.'[14]

While my situation is not in any way comparable to Kaminski's (who spent over 5 months as a political prisoner in a Polish prison), my role as committee chair and researcher is akin to an 'observing participant' as I benefit from my personal experiences of chairing and contributing to the committee's activities. However, one of the methods deployed in this paper to enable critical distance is the use of a thematic review of the extensive published committee papers and with reflection and evaluation of the committee's deliberations and the environment in which it is operating informed by relevant literature and doctrinal analysis. In addition, the self-reflection within the paper combined with the anonymous peer-review process are of themselves ways of achieving critical distance. The committee will shortly be expanding its membership and I hope that ongoing research and assessment of the committee's activities will include the opportunity for a form of 'member checking': that is, for members, who have distinctive roles and no doubt different perspectives to myself, to assess and critique my findings and conclusions.

To conclude this methodological discussion, the research described in this paper could be an example of a method that I have previously described as an 'in-house' approach to legal research. This approach is defined by practical knowledge of an operational context built up through deep interaction with that context, awareness of relevant legal frameworks, translation of theoretical concepts for such operational contexts, and the independence necessary to produce robust conclusions and recommendations from such comprehensive understanding, which then feed back into the operational context.[15] There is much pressure on academics to generate 'impact' with their work, yet considerable challenges of being a 'observing participant', and therefore I suggest that this research approach is worthy of further consideration and development within academia.

1. The contribution to effective accountability in respect of the Lab's undertakings: the six big themes

The WMP committee comprises of volunteer members (including the chair and vice-chair) with expertise in data science, law, human rights, ethics, victimisation and social exclusion and representatives of the community, and senior representatives of the Office of the Police and Crime Commissioner and WMP. The committee reviewed eighteen data analysis/modelling projects between April 2019 and September 2020. Thirteen of these projects were submitted by the Lab; four projects were from the National Data Analytics Solution (NDAS) project; and one from the Home Office. Overall (with the exception of the

---

[14] Marek M. Kaminski, *Games Prisoners Play: The Tragicomic Worlds of Polish Prison* (2004, Princeton, NJ: Princeton University Press, 2004), 7. Many thanks to the anonymous reviewer for referring me to this work.

[15] Marion Oswald, 'Cyberlaw, Interdisciplinarity and the 'In-House' Approach to Legal Research' SLSA Annual Conference 2021 https://researchportal.northumbria.ac.uk/en/activities/cyberlaw-interdisciplinarity-and-the-in-house-approach-to-legal-r.

Home Office facial recognition pilot) the projects can be categorised into two main groups based on the type of data analysis model: first, predictive models, aimed at identifying how well different factors and variables could generate particular predictions. Most often, the proposed outputs of these models were intended to influence further actions to be taken by the police, such as to implement certain supportive interventions for individuals who may be identified as likely to commit serious crimes in the future. The other group of projects were exploratory, explanatory and network analysis models where the police were interested in understanding trends, connections between people, crimes and locations, or important factors relating to an identified problem. NVivo software was used to identify and code areas of commonalities trending in the committee's discussions.[16] Identification of the themes below was based only on the published minutes of the committee which detail the discussions undertaken during the review of the proposals, including questions raised and subsequent recommendations.[17]

Addressing important policing and societal issues

Most of the projects reviewed by the committee received its commendation for the scope of the overarching problem they sought to address. Among these were:

- the 'Disproportionality' project which analysed WMP data for evidence of disproportionality, in particular by ethnicity and sex;
- against the background of the year-on-year decline in the proportion of cases resulting in a charge,[18] the 'RASSO' (Rape and Serious Sexual Offences) project which examined sexual crime data available to West Midlands Police with a view to considering the factors that influence outcomes;
- the 'Youth and Most Serious Violence' project which aimed to consider in particular whether there were particular factors that could inform why or when young people get involved in serious violence; and
- the 'County Lines' networks analysis which planned to identify links between 'nominals' (a policing term for an individual) involved in county lines drug distribution and serious organised crime.

The committee is reliant on the explanations provided by the force and the Lab in respect of the connection between the policing problem and the data analytics, and the impacts on individuals that may result, and we have not yet been able to monitor and assess the consequences of a tool in a live deployment and against actual policing 'tradecraft'. Furthermore, the committee's case-by-case approach to review, and lack of ability to

---

[17] All papers and minutes can be found at https://www.westmidlands-pcc.gov.uk/ethics-committee/ethics-committee-reports-and-minutes/.
[18] HM Government 'The end-to-end rape review report on findings and actions' June 2021, CP 437 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/994816/end-to-end-rape-review-report.pdf.

initiate investigations, does not lend itself to longer-term thematic review[19] or to the assessment of the cumulative effect of an increase in the use of data analytics, both within the police and their partner agencies, with such assessment thus falling to existing oversight, regulatory or political bodies. The committee has neither the remit nor the practical resources to review and critique the risk-based pre-emptive approach underlying many data analytics projects within the police and other public services. Redden et al. argue these approaches shift focus 'away from…underlying causes (prevention) to operationalism (pre-emption)'.[20] The 2021 Independent Review of Children's Social Care found risk-based systems resulted in assessment and investigation, not support[21] and the Independent Inquiry into Child Sexual Abuse raised a concern that 'a distinctive professional language around child sexual exploitation has developed over many years, which describes children being 'at risk' despite clear evidence of actual harm having occurred. Examples of this include children having contracted sexually transmitted diseases.'[22] Any overarching regime of accountability and oversight which lacks a system for proactive longer-term thematic review will ultimately be incomplete.

Operationalisation and subsequent interventions

The committee commonly requested further information regarding how the models would improve current processes including those involving professional judgement, how they compared to other means of achieving the same objective and how this would be evidenced and evaluated. Discussions often focused on the overall outcome that was sought from the project as a whole and whether and if so, how, data science methods could contribute positively to achieving the intended outcome. Consequently, the relationship between the output of the models and the resulting interventions were regularly the subject of questions and debate, with particular attention being paid to interventions that might be coercive or perceived by the individual as intrusive or stigmatising. (The term 'intervention' was often used in project proposals in a generic way to cover a wide variety of intentional policing involvement or interaction with a person of policing interest, and the specific details of each intervention were therefore queried and discussed.).

While predictive tools were particularly scrutinised in this respect (in the light of the non-significant precision rate of some models), network-related projects, addressing gaps in intelligence gathering, also raised concerns around the consequences of individual categorisation by such methods. It was noted that categorisations were probabilities (for instance as an individual associated with a modern slavery crime or with those involved in serious crime). The committee expressed concerns around the handling and labelling of

---

[19] Daragh Murray, Pete Fussey, Lorna McGregor and Maurice Sunkin 'Effective Oversight of Large-Scale Surveillance Activities: A Human Rights Perspective' 11 J. Nat'l Security L. & Pol'y (forthcoming 2021).
[20] Joanna Redden, Lina Dencik & Harry Warne (2020) 'Datafied child welfare services: unpacking politics, economics and power' *Policy Studies*, 41:5, 507-526, DOI: 10.1080/01442872.2020.1724928.. It is also beyond the scope of this article to discuss this aspect.
[21] 'The case for change' Independent Review of Children's Social Care, June 2021 https://childrenssocialcare.independent-review.uk/case-for-change/, 10.
[22] Independent Inquiry into Child Sexual Abuse 'Child sexual exploitation by organised networks' Investigation Report, February 2022.

the outputs (as a new piece of personal data). There was a risk of flagging individuals as potential future offenders rather than as victims, thus missing opportunities for safeguarding and support or subjecting individuals to more intrusive interventions. For instance, 'tagging' a young person as 'offender' or 'suspect' without knowledge or consideration of related incidents of domestic violence or child abuse could affect their subsequent interactions with the criminal justice system and other agencies.[23]

The committee raised the potential for the insights drawn from exploratory projects to be used for predictive purposes. For instance, in respect of the RASSO project, it was cautioned that even though this project was intended to inform or improve policing practices, the results could inadvertently encourage the prioritisation of only specific cases, that may be regarded as more likely to be successful due to the existence of certain factors. The committee requested to see commitment and assurance[24] that these projects would not be used to provide the rationale for avoiding the resourcing and investigation of challenging cases, but rather be used to augment police operations.

Data Protection Impact Assessment (DPIA) and legal advice

As set out in the committee's terms of reference, it is not its role to provide legal advice. WMP remains responsible for seeking independent specialist legal or data protection advice and conducting its own DPIAs on proposed projects before such projects are considered by the committee.[25] Therefore, the committee pays attention as part of its operating processes to the existence and content of legal advice and to the ongoing development of DPIAs. The committee noted on a number of occasions that the outputs generated by models could be significant pieces of new personal data that may influence or inform subsequent police decisions, or could be used in different contexts or by other relevant agencies, such as immigration authorities. As a result, the committee recommended that these projects must be supported by appropriate DPIAs which will address concerns about how outputs could be used in other contexts.[26] The committee emphasised that the need for DPIAs and specific legal advice was not merely a matter of legal 'compliance' but to guide the police and the committee through legal and ethical issues that may be suspected, perceived or identified in proposed projects. Furthermore, legal advice was critical to the assessment of the necessity and proportionality of using certain personal data, such as ethnicity, gender, drug use and age in predictive and network analysis models, and for considering the implications of using and sharing the outcomes. I

---

[23] Murphy, Daniel S., Brian Fuleihan, Stephan C. Richards, and Richard S Jones. 2011. "The Electronic "Scarlet Letter": Criminal Backgrounding and a Perpetual Spoiled Identity." *Journal of Offender Rehabilitation* 50 (3): 101–118.

[24] Committee minutes July 2019 https://www.westmidlands-pcc.gov.uk/ethics-committee/ethics-committee-reports-and-minutes/.

[25] Terms of Reference, para 5 https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/07/Ethics-Committee-Terms-of-Reference-as-at-1-April-2019.pdf?x39505.

[26] The Metropolitan Police's gangs database is one well known example of the potential harms resulting from the 'spread' of data-generated categorisations: Vikram Dodd 'Met removes hundreds from gangs matrix after breaking data laws' *The Guardian* 15 February 2020 https://www.theguardian.com/uk-news/2020/feb/15/met-removes-hundreds-from-gangs-matrix-after-breaking-data-laws.

reflect further on the relationship between ethical and legal advice in sections 2 and 3 below.

Data inputs and the validity of outputs

Where a model was intended for prediction purposes, this raised questions both about the appropriateness and reliability of the underlying variables of the model and the overall predictive power of the model. Concerns were often raised over the limitations of the data sets that were relied on for prediction, exploratory or network analysis purposes, and whether the data available represented all the relevant information that should be taken into account to produce a reliable outcome, such as to give a full picture of an individual who may benefit from an intervention. The Community Tension project, for instance, was withdrawn after the Lab concluded that it had not proved possible to satisfactorily link reports of community tension to incidents of violence (or other forms of crime) and therefore no predictive model could be built. This project was therefore discontinued by the Lab, with the Lab's critical approach being commended by the committee.[27]

The committee's recommendation that the 'Most Serious Violence' model, proposed by the National Data Analytics Solution project,[28] did not proceed was based primarily on the model's lack of statistical validity (with precision rates of 38% at best)[29] and lack of sufficient safeguards around coercive interventions that could result.[30]

For some exploratory models, such as understanding disproportionality, the committee advised that these projects could have potentially strong benefits only if they considered other factors which influence discriminatory and other practices. However, it was observed that these relevant contextual factors were absent from the police data, reflecting common concerns in academic literature regarding the validity of individual predictive analytics.[31]

The committee advised that it was important that the probabilistic nature of outputs of data analytics was recognised, and that outputs were handled appropriately as a form of intelligence in accordance with the standard '3-5-2' process of grading intelligence[32] thus

---

[27] WMP Data Analytics Lab 'Community Tension' paper, July 2020 https://www.westmidlands-pcc.gov.uk/ethics-committee/ethics-committee-reports-and-minutes/.

[28] https://west-midlands.police.uk/about-us/privacy-notice/national-data-analytics-solution.

[29] National Data Analytics Solution submission to the WMP Ethics Committee July 2020 file:///C:/Users/vgyk3/Downloads/07072020-EC-Agenda-Item-9-NDAS-Update.pdf

[30] Ethics Committee minutes 1 July 2020 file:///C:/Users/vgyk3/Downloads/01072020-EC-Minutes-and-Advice%20(1).pdf.

[31] See for instance Matthew J Salganik et al., 'Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration', *Proceedings of the National Academy of Sciences* (Vol. 117, No. 15, 2020), pp. 8398–403.

[32] Ministry of Justice 'Intelligence Collection, Analysis and Dissemination Policy Framework' 31 October 2019 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/843803/intelligence-collection-management-dissemination-pf-31-oct-2019.pdf.

ensuring that uncertainties and confidence ratings around the outputs were highlighted and communicated when data were further disseminated.[33]

'Biases' in data, false positives and safeguards

Linked to the above concern around validity of outputs were issues of inaccurate categorisation, such as level of risk of transitioning to commit serious offences or links to a category of crime such as modern slavery. This issue was stressed in all the geographically based models which sought to predict hotspots of crime and other models which had ethnicity data (or potential proxy data) as input. The committee commented upon the risk of stigmatising different communities, perpetuating ethnic stereotypes, or confirming historic disproportionate actions.

To address potential biases and false positives, the committee emphasised that models should not be operationalised without having first determined an evaluation process and criteria which will be used to identify false positives and false negatives, and to assess potential consequences should erroneous outcomes be acted upon. Such evaluation should consider the way the outputs of models will be understood, how they will be communicated to officers and their impact on decision-making, the purpose for which they will be used, who will have access to the information, where outputs will be stored, and whether their use will be limited to non-coercive interventions.

Transparency and Partnerships

With the possibility of data sharing between the police and other agencies, the committee often raised concerns for vulnerable victims and individuals, and whether limitations should be placed on communication with other agencies, such as immigration authorities.[34] The committee often advised on the need for a clear communication strategy which should inform police officers and members of the public as to how the findings of projects will be put to use, and on the need for increased clarity between the police and other public agencies which may be recipients of information from data analytics, on the intended operational uses and the meaning of the outputs.

The penultimate brief sub-section below comments upon the committee's purpose, function, structure and operation which link to the long-term viability of the approach.

Status, function and operation – challenges and opportunities

Scholars have emphasised the importance of considering 'data systems as complex assemblages of artefacts, people, programmes, infrastructures, ideas, etc' and for evaluation of their social and technical impact to take greater account of context, design

---

[33] n4, see recommendation on page x.

[34] These concerns reflect recent debate in the House of Lords over the Domestic Abuse Bill and amendments tabled to prevent personal data of a domestic abuse victim accessing government support being processed for an immigration control purpose: Hattie Williams 'Lords approve Bishop's amendment to Domestic Abuse Bill to protect migrant women' Church Times, 18 March 2021 https://www.churchtimes.co.uk/articles/2021/19-march/news/uk/lords-approve-bishop-s-amendment-to-domestic-abuse-bill-to-protect-migrant-women.

and application.[35]   The committee has made progress towards treating data analytics within policing as 'data assemblages'[36] and thus paying attention to the realities on the ground and potential impacts on rights.

However, although the committee was founded by the Police and Crime Commissioner who has a statutorily defined role to hold the Chief Constable to account, the committee itself has no statutory or formal regulatory function, no independent existence, nor powers to require information, commission research, initiate specific or thematic investigations or to insist that its advice is implemented, and therefore may have limited ability to assess or influence relevant structural factors and long-term consequences of AI.[37] The committee's very existence is precarious, and subject to political, funding and resourcing changes, and cannot in its current form be a replacement for formal oversight. The terms of reference were inspired to a large extent by the National Statistician's Data Ethics Advisory Committee,[38] but within policing, its approach has been developed somewhat in isolation. Despite these risks, the committee has managed to provide additional and transparent positive pressure on the force to improve – *productive challenge* - as well as a certain level of validation for the projects that have progressed.

It may have initially seemed strange to some within policing that the committee's proceedings focused upon operational outcomes – how to 'do policing' – rather than pure technology/data. The link between the two, and in particular the need to understand whether the tool in its operational context is about 'care or control' (and thus how its consequences will affect individuals) is now better understood. The committee is tasked in the terms of reference with providing 'pragmatic advice that is appropriate for meaningfully advising the CC and PCC on how to move forward,'[39] requiring an understanding of fast-moving operational requirements and the implications of *not* doing data analytics. The committee's first three years have obliged both the committee members and those presenting their projects to find a way to operate constructively, to navigate cultural and disciplinary clashes, and to generate understanding of different perspectives and motivations. Constant attention will be needed to ensure that the committee continues to provide 'points of friction' in the scrutiny process and avoids unwarranted deference or association with police forces.[40]

The diverse skills, experiences and backgrounds of its volunteer members probably represent its biggest strength, but also one of the biggest challenges in terms of obtaining

---

[35] n20.
[36] Kitchin, Rob, and Tracey Lauriault. 2014. "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work." The Programmable City Working Paper 2. http://ssrn.com/abstract=2474112.
[37] Matthew Le Bui and Safiya Umoja Noble, 'We're Missing a Moral Framework of Justice in Artificial Intelligence: On the Limits, Failings, and Ethics of Fairness' in *The Oxford Handbook of Ethics of AI*, eds. Markus D. Dubber, Frank Pasquale and Sunit Das (Oxford 2020).
[38] https://uksa.statisticsauthority.gov.uk/about-the-authority/committees/national-statisticians-data-ethics-advisory-committee/.
[39] Terms of Reference, para 47 https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/07/Ethics-Committee-Terms-of-Reference-as-at-1-April-2019.pdf?x39505
[40] n19.

similar levels of diverse expertise should such a model be rolled out regionally or nationally. Recent self-assessment raised the need to further represent the voices of young people and those with lived experiences in committee discussions,[41] a gap that we have attempted to fill in the most recent recruitment exercise. The model could of itself be developed within existing ethics or other advisory structures, although the need for diversity of relevant skills, expertise and backgrounds would remain the same.

Finally in this section, it is worth pointing out again that the committee has an advisory function in respect of Lab projects (although it has with agreement reviewed a national policing project and a Home Office proposal). It can only review what is given to it. It has no access to an audit function in respect of how its recommendations are implemented or communicated. A national approach based on this model should clarify the criteria on which projects would be triaged for review, the rolling nature of review (from idea to operational) and how implementation of committee recommendations would be audited.

2. Law vs ethics, or law and ethics

'Ethics is *both more and less* than law: it is more because many ethical concerns are not addressed by the law and less because the outcome of ethical considerations are not necessarily transformed into legal norms and thus not enforceable by way of law.'[42] Yeung et al. criticise the 'vagueness and elasticity' of AI ethics principles, operating as 'an empty vessel into which anyone…can pour their preferred ethical content.'[43] Such critiques highlight the challenge faced by members of so-called 'ethics' bodies – how to balance opportunities for ethical reflection with the need for closure and operational certainty, and how to ensure 'social acceptability' while not 'crippling the chances to harness the social value of data science'[44] by over-rigid focus on individual rights.

There is no shortage of ethical principles proposed for artificial intelligence (AI) and data analytics. Fjeld at al.'s comparison of thirty-six 'fractured' sets of AI principles highlighted certain coalescing topics: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values.[45] Such bland, rather self-evident themes are hardly likely to have real teeth. Impact will depend upon the process of embedding 'in a larger governance ecosystem, including for instance relevant policies…, laws, regulations, but also professional practices and everyday routines'[46] and most high-level principles have yet to be translated into contextual policy, operational guidance and professional practice.[47] Frameworks which tend to focus only upon 'data in' and project methodology

---

[41] Minutes of the committee 5 March 2021 https://www.westmidlands-pcc.gov.uk/ethics-committee/ethics-committee-reports-and-minutes/.

[42] Mireille Hildebrandt *Law for Computer Scientists (*Oxford 2019).

[43] n10.

[44] n42.

[45] Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI." Berkman Klein Center for Internet & Society, 2020.

[46] n45, 5.

[47] n4.

in an attempt to make it 'more ethical' (such as the UK Data Ethics Framework[48]) potentially obviate personal responsibility for assessing the validity and wider consequences of AI deployment objectively using all relevant evidence and factors. Furthermore, conflating ethics with fundamental rights can 'reduce[s] human rights to very basic principles such as "do good" or "do no harm" which are extraordinarily far removed from actual fundamental rights.'[49]

The purpose of this paper is not to engage in 'ethics bashing'[50] however. Ethical deliberation has an important role as part of a wider approach to achieving trustworthy use of emerging technology within policing, and in particular to broadening the context of the debate. It is unfortunate that, as Bietti highlights, 'ethics' has become associated with 'objectionable instances of self-regulation, static and incomplete lists of guiding principles and other forms of narrow and conservative regulative "fixes"'.[51] In light of the departure from Google of two leading members of its 'Ethical AI' team, Naughton criticises the 'ethics theatre' of 'ethics boards, panels and oversight bodies established by the same companies' and the involvement of 'entrepreneurial academics anxious to get a slice of the action' with 'lucrative consultancies to advise on ethical issues raised by machine learning…becom[ing] a vast system of out-relief for otherwise unemployable philosophers and other sages.'[52]

Arguably the committee, as a form of self-regulation, could be deserving of similar criticism. However, the committee, despite being badged with the ethics label, performs a wider independent advisory function, contributing to consideration of questions of scientific validity, legal proportionality and operational context. The principles against which projects are reviewed include necessity and proportionality, human oversight of the limits and risks of new technologies, analysis methods that are sufficiently accurate, appropriate and rigorous in order to draw reasonable conclusions, and non-stigmatisation.[53] These principles reflect in part Yeung et al.'s call for governance models that are anchored in a human rights approach,[54] and Murray et al.'s emphasis on processes that include critical adversarial voices that can challenge existing assumptions and bring in different perspectives.[55] It is inadvisable to separate ethics and standards from relevant

---

[48] Government Digital Service 'Data Ethics Framework' (Updated 16 September 2020) https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020.
[49] Sylvie Delacroix and Ben Wagner, 'Constructing a mutually supportive interface between ethics and regulation' (2021) *Computer Law & Security Review* 40.
[50] Elettra Bietti, From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy (December 1, 2019). DRAFT - Final Paper Published in the Proceedings to ACM FAT* Conference (FAT* 2020), Available at SSRN: https://ssrn.com/abstract=3513182.
[51] n50.
[52] John Naughton 'Google might ask questions about AI ethics, but it doesn't want answers' *The Guardian*, 13 March 2021 https://www.theguardian.com/commentisfree/2021/mar/13/google-questions-about-artificial-intelligence-ethics-doesnt-want-answers-gebru-mitchell-parrots-language.
[53] Terms of Reference, para 48 https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/07/Ethics-Committee-Terms-of-Reference-as-at-1-April-2019.pdf?x39505.
[54] n10.
[55] n19.

law and the wider operational context.[56] Collective deliberation (demonstrated by the 'big themes' discussed above) and professional ethical standards as a 'dynamic system'[57] can bring to the surface the serious and messy challenges faced within policing, acting as one important pillar supporting the aim of achieving trustworthy use of emerging technologies in policing. Individual-centric perspectives within the law and litigation, and *ex post* forms of evaluation and governance can obscure or overlook the longer-term consequences of AI, such as the gradual impact on group rights**,** the implications of AI feeding AI within critical systems and the consequences for the position of the individual police officer as responsible for their own professional and ethical behaviour.[58] As Bietti argues, '[e]thical reasoning or moral inquiry can have intrinsic value as a process and instrumental value as a means to the achievement of other valuable outcomes.'[59]

However, the committee's approach is primarily a human rights focused one, with its considerations reflecting the key factors of the human rights necessity test. The application of the necessity and proportionality tests, combined with laws relating to specific policing activity and the development of clear scientific standards, should help us to avoid falling off the 'ethics edge…where the law no longer guides and protects us, leaving ethics as the lone standard by which to gauge our behaviour.'[60]

3.   A human rights-based approach

Based on qualitative interview data of operational police uses of advanced surveillance technologies, Fussey and Sandhu point to how

> 'recourse to subjective ethical codes *and* a more general, utilitarian, 'public good' are judged sufficient grounds for licentious surveillance practices and dismissal of legal safeguards. Such justifications hold multiple implications. Particularly prominent is the instrumental and ironic way personal 'ethics' are used as justification that unethical extra-procedural practices are somehow ethical.'[61]

Fussey and Sandhu also highlight the tendency to view public safety as the 'higher purpose' over citizen's rights, misconstruction of the standard human rights legal test and the characterisation of proportionality 'in terms of achieving policing aims rather than addressing the degree of rights interference.'[62]

---

[56] n4: This is reflected in the recommendations of Babuta and Oswald (2020) for an integrated impact assessment for data analytics within policing, to be reflected in HMICFRS's inspection regime.
[57] n49.
[58] College of Policing: Code of Ethics, July 2014, para 1.4.1 https://assets.college.police.uk/s3fs-public/2021-02/code_of_ethics.pdf.
[59] n50.
[60] Susan Liautaud *The Power of Ethics* (Simon & Schuster 2021) 3.
[61] Peter Fussey and Ajay Sandhu. 'Surveillance Arbitration in the Era of Digital Policing.' *Theoretical Criminology*, (October 2020). https://doi.org/10.1177/1362480620967020.
[62] n61.

Where applied appropriately however, McGregor et al. argue that the application of a human rights framework 'is intended to ensure that the potential inherent in technology can be realized, while at the same time ensuring that technological developments serve society.'[63] Accountability proposals should 'be set within a wider framework, addressing the overall algorithmic life cycle, from the conception and design phase, to actual deployment and use of algorithms in decision-making.'[64]

Such wider framework should assess whether data analytics can help forces meet their article 2 and 3 positive obligations. In arguing for more investment to explore the use of data-driven technology to predict, prevent and pursue criminal harms against women, Grace suggests that the judgment in the case of

> '…DSD[65] means that the police must be proactive in pursuing serial offenders displaying violent criminality in their behaviour if they are to meet their Article 3 ECHR obligations, and this is a consideration that is to some degree disconnected from the question as to whether a conviction is eventually arrived at. The question then arises, in the context of growth in the sophistication of machine learning informed policing, is whether data-driven tech can help police forces meet their Article 3 ECHR investigative duties in particular, following DSD, as well as their wider obligations under Articles 2 and 8 ECHR in addition.'[66]

A criticism that could be levelled at a human-rights based approach to policing and technology is that it lacks a corresponding regulatory infrastructure.[67] According to Murray, 'using human rights law to inform states' decision-making processes is not straightforward. Although human rights law imposes (essential) ex ante obligations, our understanding of how that law applies, and the content of specific obligations in specific contexts, is primarily derived from ex post accountability mechanisms. These understandings do not apply straightforwardly to ex ante processes.'[68]

Operationalising the necessity test

The proceedings of the committee arguably contribute to addressing the challenges identified above and in particular the practical application of factors within the necessity test to new technological developments in policing.[69] As Murray points out, 'if states must ensure that their activities do not result in human rights violations, they must identify the

---

[63] McGregor, L., Murray, D., & Ng, V. (2019) 'International Human Rights Law as a Framework for Algorithmic Accountability' *International and Comparative Law Quarterly, 68*(2), 309-343.

[64] n63.

[65] *Commissioner of Police of the Metropolis v DSD and another* [2018] UKSC 11. Lord Kerr at para. 29.

[66] Grace, Jamie, Female victims of gendered violence, their human rights and the innovative use of data technology to predict, prevent and pursue harms (January 29, 2021). Available at SSRN: https://ssrn.com/abstract=.

[67] n11, 36.

[68] Murray, D. (2020). Using Human Rights Law to Inform States' Decisions to Deploy AI. *AJIL Unbound, 114*, 158-162.

[69] The ECHR requires the measure in question: (a) to be in accordance with the law, (b) to pursue a legitimate aim, and (c) be necessary in a democratic society. This paper will focus upon 'necessary in a democratic society.'

potential impact of those activities.'[70]   The majority of the projects reviewed by the committee engage ECHR rights, not only Articles 5 (liberty and security), 6 (fair trial), 8 (respect for private life) and 10 (freedom of expression), but also Articles 2 (life) and 3 (freedom from torture and inhuman and degrading treatment) in relation to positive protective or preventative obligations.[71] Data analytics projects that aim to bring together sources of data to determine harm or risk, or identify previously-hidden cases of serious crime, although likely based upon the police's preventive and public protection duties, will raise rights issues relating to the inputs used and deployment of the algorithmic categorisation.

The test typically used by the courts is laid out in four parts: (a) is the objective *sufficiently important* to justify limiting a fundamental right (a pressing social need)? (b) are the measures which have been designed to meet it *rationally connected* to it? (c) are they *no more than necessary* to accomplish it? (is a less intrusive measure available?) and (d) do they strike *a fair balance* between the rights of the individual and the interests of the community?[72] Murray distils this test into two main criteria: why is a deployment required, what alternative mechanisms are available?[73]

As demonstrated by the 'big themes' discussed above, the committee's primary focus is the policing purpose of the project at a fairly granular level and the way that data or technology will be used to contribute to that purpose. In general, projects were linked to crimes and societal issues of a serious nature likely to hit the threshold of a pressing social need. The committee's enquiries however focused on exploring the specific issue to be addressed, the strength of the connection of the data analytics to that issue, and how the proposed data analytics would be deployed in practice: 'This is essential to evaluating impact—both in terms of utility and harm'.[74] The committee often raised the question of how the proposed method compared to existing processes and if not yet known, how this would be evaluated, crucial considerations to the evaluation of proportionality.

Reflecting McGregor et al.'s recommended 'lifecycle' accountability approach, the committee's terms of reference anticipate a rolling rather than a one-off process of review and advice – the committee will be given a chance to consider: a) the proposed policies and/or operational strategies being considered for interventions arising from the findings of the analysis; and b) whether or not any actioned policies and/or operational strategies were successful at achieving the stated objectives and/or an assessment of the public benefit or harm caused by such.[75] This process recognises that some AI 'tools are so new that the resource benefits have yet to be realised, and it may be too early to judge the

---

[70] n68.
[71] *Osman v UK* (1998) *(87/1997/871/1083); Opuz v Turkey (2009) 33401/02.*
[72] Lord Wilson in *R. (on the application of Aguilar Quila) v Secretary of State for the Home Department* [2011] UKSC 45.
[73] n68.
[74] n68.
[75] Terms of Reference, para 37 https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2019/07/Ethics-Committee-Terms-of-Reference-as-at-1-April-2019.pdf?x39505.

benefits and harms with ease'[76] and therefore, an iterative approach is needed. Oswald and Grace's concept of 'experimental proportionality' is designed to permit the use of unproven data technologies in the public sector in order that benefits and harms can be fully explored (provided a baseline connection to a legitimate aim and a reasonable belief that there is no excessive cost to human rights can be demonstrated), subject to meaningful and periodic review via a dedicated mechanism.[77] I would argue that the committee's proceedings to date demonstrate the potential of such an experimental proportionality approach, provided that the review mechanism is suitably robust, influential, transparent and independent, with its time and funding protected. As many of the projects reviewed by the committee are only now reaching an operational stage, however, it is yet to be seen whether the committee will be able to support and underpin a suitably robust, critical and ongoing evaluation process.

In accordance with law

The necessity test is of course only one (albeit important) aspect of the required human rights assessment. The technological measure that interferes with fundamental rights must have a basis in law that is foreseeable and accessible.[78] As the Court of Appeal confirmed in *Bridges*, case-law of the European Court of Human Rights has established that local policies can contribute to satisfying the 'accordance with law' requirement, provided such policies give sufficient clarity as to how the public body's discretion would be exercised to protect against arbitrary interference with human rights.[79] The Court of Appeal accepted a 'relativist approach' to the precision of the law used to justify an intrusion: the more intrusive the method, the more specific and precise the law must be to justify it.[80]

It is beyond the scope of this paper to consider whether the law governing police use of different methods of data analytics is sufficiently precise. The committee does represent a form of 'self-restraint'[81] implemented by the police force itself. It is questionable however whether the operations of the committee to date would of themselves remedy any lack of precision in the law. The committee proceedings, although subject to detailed terms of reference, do not equate to a 'policy' in the sense of determining with precision the terms on which a data analytics tool would be used (although may well contribute to the proportionality requirement). The proceedings and determinations could, however,

---

[76] Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes, 'Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality' (2018) *Information & Communications Technology Law*, 27:2, 223-250, 242.

[77] n76, 34.

[78] *Malone v United Kingdom* (1984) 7 EHRR 14.

[79] *R (on the application of Edward Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058, para 61.

[80] n79, para 83.

[81] *Beghal v DPP* [2015] UKSC 49 per Lord Kerr [102].

contribute to the development of appropriate policy and therefore to a framework that could be regarded as having the necessary quality of law.

4. Law governing policing activities

In their examination of the AI regulatory agenda, Black and Murray comment on the limits of general frameworks such as human rights:

> 'In areas where AI is being used where there is currently no regulation or it falls at the edges of existing regimes, then we will have to rely on existing legal principles. Reed, for example, argues the application of general legal principles, in particular human rights, can provide an interim framework for the general regulation of AI.[82] But there are limits to the degree to which general legal frameworks, such as the law of negligence, may adequately be used to manage risks or attribute liability in ways which achieve overall societal goals.'[83]

Policing in England and Wales, as in many jurisdictions, is subject to a plethora of specific legislation, statutory codes and authorised professional practice, with human rights - Articles 5, 6 and 8 in particular - at their heart, and it is therefore surprising that relatively little attention has been paid to these laws and codes in the context of AI-informed police decision-making. For instance, 'reasonable grounds for suspicion' are in most cases[84] required for the exercise of stop and search powers. Code A pursuant to the Police and Criminal Evidence Act 1984 attempts to define the concept.[85] Stop and search must be used 'fairly' and without unlawful discrimination.[86] The public sector equality duty[87] applies to stop and search powers and therefore a decision to stop and search must not be based on protected characteristics. Code A emphasises that a decision must be based on 'objective' factors,[88] as explained by Lord Justice Purchas in *Castorina*:

> 'suspicion must arise from reasonable cause. Reasonable cause, it is not disputed, is to be determined as an objective matter from the information available to the arresting officer and cannot have anything to do with the subjective state of the officer's mind.'[89]

Personal factors, either alone or in combination, cannot support reasonable suspicion. Such factors include physical appearance, the fact of previous convictions, and

---

[82] Chris Reed, 'How should we regulate artificial intelligence?' *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* , 376 (2128) 20170360, 2018.

[83] Julia Black and Andrew Murray 'Regulating AI and Machine Learning: Setting the Regulatory Agenda' *EJLT* Vol. 10 No. 3 (2019).

[84] 'No suspicion' search powers are set out in section 60 Criminal Justice and Public Order Act 1998.

[85] PACE Code A 2015: Code of practice for statutory powers of stop and search and requirements to record public encounters by police officers and staff
https://www.gov.uk/government/publications/pace-code-a-2015.

[86] n85, para 1.1.

[87] Equality Act 2010, section 149.

[88] n85, para 2.8A.

[89] *Castorina v. Chief Constable of Surrey (1988) N.L.J. 180, 181.*

'generalisations or stereotypical images that certain groups or categories of people are more likely to be involved in criminal activity.'[90] In *Black*, the appellant had arrived at his brother's house, which was being lawfully searched by the police for drugs. He was immediately detained by the police on the basis that 'he was visiting a well-known drug dealer' and he might be there to buy or sell drugs. The Divisional Court rejected this as grounds for reasonable suspicion. This was a generalised suspicion, not a specific one:

> 'reasonable grounds for suspecting could not be said to exist in this case if any person was obviously there in some lawful capacity; a visitor who called for a catalogue payment, or a door-to-door salesman, or the gas man, to take some obvious examples. The appellant was a known close relative of the occupant. He therefore had an independent reason to be there, quite apart from any suspicion relating to drugs.' (Mitchell J) [91]

What can we take from these cases in terms of the use of the outputs of AI to inform police decision-making? First, that we too should not make generalisations. We must understand the specific decision-making processes and related legal standards into which the algorithmic output is being introduced, and we must understand what the tool *does* (rather than relying on vague 'AI' descriptions that might tempt the individual to overly defer to or mythologise the output). I have previously argued that

> 'the forecasts produced by many existing algorithmic tools are probabilities (that the person or situation in question has a certain similarity to people or situations in the past). But they appear at times to be presented as something more: a prediction of reoffending becomes a 'risk' of reoffending and thus the risk if, say, a person is given parole. Determinations of risk—a decision for the public body—may depend upon many considerations, including what is unknown and the impact of the thing that is predicted. The point at which that determination is made, however, could inadvertently be moved back to the model-creators by the way that outputs are presented.'[92]

This observation mirrors Barabas's concern with the conflation of algorithmic outputs with conclusions to constitutionally required legal tests, despite 'data used to measure accuracy [not being] representative of the outcome of interest.'[93]

Algorithmic forecasts are no more than 'good bets'[94] or in the words of the Supreme Court of Wisconsin in the famous *Loomis* case:

---

[90] n85, para 2.2B.

[91] *Michael David Black v Director of Public Prosecutions* CO 877-95 High Court of Justice Queen's Bench Division (Divisional Court) 1995 WL 1083760.

[92] Marion Oswald 'Algorithmic-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power' (2018) *Phil. Trans. R. Soc. A,* 376:2128.

[93] Chelsea Barabas 'Beyond Bias: "Ethical AI" in Criminal Law' in *The Oxford Handbook of Ethics of AI*, eds. Markus D. Dubber, Frank Pasquale and Sunit Das (Oxford 2020).

[94] Berk RA, Bleich J. (2013) 'Statistical procedures for forecasting criminal behaviour: a comparative assessment' *Criminology & Public Policy* 12, 513–544.

'risk scores are intended to predict the general likelihood that those with a similar history of offending are either less likely or more likely to commit another crime following release from custody . . . the risk assessment does not predict the specific likelihood that an individual offender will reoffend. Instead, it provides a prediction based on a comparison of information about the individual to a similar data group.'[95]

Outputs should not be presented as more than probabilities (that the person or situation in question has a certain similarity to people or situations in the past) if they are to be used appropriately. Probabilistic outputs of risk or harm based on comparison with other people in the past cannot, I would argue, satisfy the requirement for reasonable grounds, as they would fall within the exclusions of generalisations, category-based suspicion, and suspicion based on general association as in *Black* (not to mention concerns around false positives). Wider concerns about fairness and dignity may explain the committee's regular scrutiny of the data sets that were relied on for prediction, exploratory or network analysis purposes, and related concern around whether an individual would be assessed *as an individual*. This may reflect an underlying awareness that, as Hannah Fry puts it, '[t]o simplify the world enough that it can be captured with numbers means throwing away a lot of detail.'[96]

That is not to say that algorithmic analysis has no role to play in policing, quite the reverse as the committee has indicated on a number of occasions. Pasquale argues that AI could advance the human condition if it complements professionals, rather than substituting them.[97] Where previously siloed data is brought together, the results can reveal a bigger picture about an individual or circumstance, one that might indicate the need for urgent safeguarding, proportionate preventative or public protection action,[98] potential indicators of causation, strategic planning and resource deployment, or avenues for further investigation. But while the analysis may give us a bigger picture, it will not give us the full picture – it provides us with intelligence which may justify certain policing activities but not others. Algorithmic fairness cannot be narrow in scope (limited to fine-tuning models or datasets); a 'fundamentally equitable path must examine the wider picture', prioritising 'in-depth and contextual understanding.'[99] In order to decide what we should do with the output, we need to know what it is telling us (and not telling us), and this is where the science comes in.

---

[95] *State of Wisconsin v Eric L. Loomis*, 2016 WI 68.

[96] Hannah Fry 'What data can't do' *The New Yorker*, March 29, 2021
https://www.newyorker.com/magazine/2021/03/29/what-data-cant-do.

[97] Pasquale, F. *New Laws of Robotics: Defending Human Expertise in the Age of AI* (Harvard University Press, Massachusetts, 2021).

[98] As Jamie Grace has highlighted in his work around the Domestic Violence Disclosure Scheme (Clare's Law) n66.

[99] Birhane, A. 'Algorithmic injustice: a relational ethics approach' (2021) *Patterns*, 2(2).
https://doi.org/10.1016/j.patter.2021.100205.

5.   The importance of scientific validity

Bearing in mind its multidisciplinary membership, including both academic and commercial data scientists, it was unsurprising that the committee should pay attention to the statistical and scientific validity of the proposed methods, and to the assumptions and values built into the analysis. The answer to the basic question 'does it work?' is of course fundamental to the necessity assessment as discussed above. It is beyond the scope of this paper to adequately summarise the extensive debates around the validity, or otherwise, of various data science methods and AI applications, and in particular predictive methods at an individual level. However, my research relating to AI and national security concluded:

> 'the evidence reviewed for this paper suggests that it is neither feasible nor desirable to attempt to develop AI systems to 'predict' human behaviour at the individual level – for instance, for counterterrorism risk assessment purposes. Nevertheless, [Augmented Intelligence] – the use of AI systems to collate relevant information from multiple sources and flag significant data items for human review – has clear potential benefits in this context and is likely to improve the efficiency of analysis tasks focusing on individual subjects.'[100]

These concerns are ongoing. Salganik et al.'s 2020 research into prediction of life outcomes using a mass collaboration approach concluded that 'despite using a rich dataset and applying machine-learning methods optimized for prediction, the best predictions were not very accurate and were only slightly better than those from a simple benchmark model'.[101] Lum et al.'s analysis of variation in individual-level probabilities of failing to appear for court (available in a 2021 pre-print) focused upon statistical confidence in labels generated by actuarial risk assessment instruments:

> 'when individual propensity toward the outcome that is not explained by covariates is explicitly modeled, there is significant uncertainty about the probability of the outcome for most individuals. As a result, there is also significant uncertainty about the risk group to which most individuals belong: the between-individual variability swamps what little signal there is in the covariates.'[102]

These technical issues cannot be 'left' to the data scientists to address. This is because AI and data analytics create labels – or data shadows as they could be called – that attach themselves to individuals and which can affect their future treatment and life-chances in a positive or negative way. As I recommended in 2020, we urgently need 'context-specific evaluation methodologies for statistical algorithms used by police forces in England and Wales. This should include guidance on how confidence levels and error rates should be

---

[100] Alexander Babuta, Marion Oswald and Ardi Janjeva (2020) 'Artificial Intelligence and UK National Security: Policy Considerations' *RUSI Occasional Paper*, 16.

[101] Matthew J Salganik et al., 'Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration', *Proceedings of the National Academy of Sciences* (Vol. 117, No. 15, 2020), pp. 8398–403.

[102] Kristian Lum, David B. Dunson and James Johndrow (2021) 'Closer than they appear: A Bayesian perspective on individual-level heterogeneity in risk assessment' <u>arXiv:2102.01135v1</u> [stat.AP].

established, communicated and evaluated.'[103] In other words, we need to know if the AI is any good for the specific police context in which it will be deployed.

It continues to be the case that 'the development of policing algorithms is often not underpinned by a robust empirical evidence base regarding their claimed benefits, scientific validity or cost effectiveness' with claims of 'predictive accuracy' often misunderstood or misinterpreted, raising difficulties around the assessment of a tool's real-world benefits.[104] Police forces in England and Wales are statutorily independent and nationally consistent minimum standards have yet to emerge. As mentioned above, the initial 'Most Serious Violence' predictive model, proposed by the National Data Analytics Solution project was withdrawn due to concerns with statistical validity. Other forces continue to investigate violent recidivism predictive models,[105] such as the 'OxRec' model developed by Oxford University, at the date of writing under evaluation by Thames Valley Police. The OxRec project provides an interface for the calculation of individual risk levels,[106] incorporating factors such as immigration status and neighbourhood deprivation that raise issues of both relevance and equalities pursuant to the public sector quality duty. The statistical validity, and in particular the high false positive rate[107] and stated positive predictive value of between 21-37%[108] also raise concerns. The authors themselves admit that 'the tool has low predictive accuracy at the individual level….Therefore, one potential harm that is not justified is preventive detention.'[109] Whether other forms of individualised police 'intervention' should be based on this tool's output is questionable. There appears to be an urgent requirement to address consistency around deployment of algorithmic models of limited accuracy, taking into account the partial, incomplete and limited nature of police-held data.[110] It would be all too easy, for instance, for algorithmic categorisations to be based purely on 'unreliable' intelligence and non-crime data in a way that was opaque to the human recipient of the algorithmic output. The appointment of a new Chief Scientific Advisor for Policing provides opportunity for progress in this regard.

6.   A three-pillar approach

The above analysis has shown that the technical, statistical, legal, contextual, operational and ethical aspects of algorithm-informed policing are closely interconnected. We need to know what the output means in the context of the operational decision to be taken, and

---

[103] n4, xi.
[104] n4, vii-viii.
[105] College of Policing 'A validation study of OxRec model for assessing risk of violent recidivism' https://whatworks.college.police.uk/Research/Research-Map/Pages/ResearchProject.aspx?projectid=922.
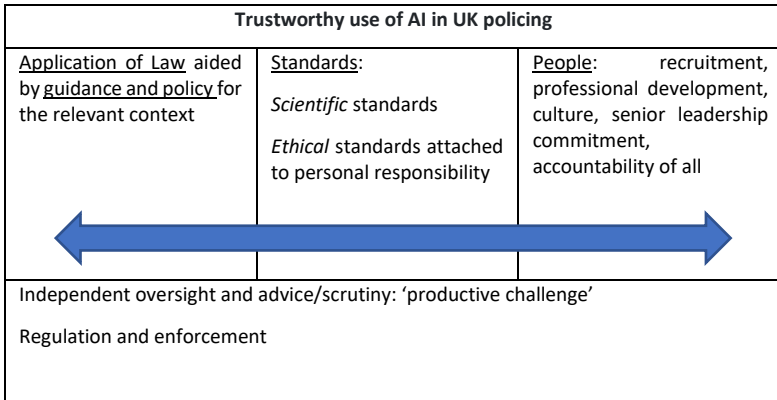[106] https://oxrisk.com/oxrec/.
[107] https://oxrisk.com/oxrec-background/.
[108] Seena Fazel et al. (2016) 'Prediction of violent reoffending on release from prison: derivation and external validation of a scalable tool' *The Lancet* 3(6) 535-543 https://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(16)00103-6/fulltext.
[109] n108.
[110] n2, 77.

we must assess the implications of how it will be used in practice. The three-pillar approach set out in *Fig. 1* attempts to illustrate the inter-linking of all these factors.

Trustworthy use of AI in UK policing		
<u>Application of Law</u> aided by <u>guidance and policy</u> for the relevant context	<u>Standards</u>: *Scientific* standards *Ethical* standards attached to personal responsibility	<u>People</u>: recruitment, professional development, culture, senior leadership commitment, accountability of all
Independent oversight and advice/scrutiny: 'productive challenge' Regulation and enforcement		

Fig. 1 Three-pillar approach to achieving trustworthy use of AI in UK policing

The **application of relevant law**, as discussed above, is not easy as it often requires the 'read-across' of common law principles into new contexts.[111] This is why robust legal advice is so important: advice that does not automatically equate necessity with a project that is pursuing policing purposes. Also crucial is the translation of key legal requirements into suitably precise policy and guidance as the *Bridges* case discussed above has emphasised.

Oswald and Grace's 'Algo-care'[112] is a decision-making framework that attempts to do just that. Each word in the mnemonic – *Advisory; Lawful; Granularity; Ownership; Challengeable; Accuracy; Responsible; Explainable* – is supplemented by questions and considerations representing key legal considerations (such as necessity and proportionality, natural justice and procedural fairness), as well as practical concerns such as intellectual property ownership and the availability of an 'expert witness' to the tool's functionality. The current Algo-care questions are incorporated into the committee submission process. The framework was designed to be used in parallel with data protection and equality impact assessments, and since Algo-care's design, the Information Commissioner's Office has produced a toolkit for police forces using data analytics which is designed to take police staff through the data protection points that they should consider from the start of any project.[113] Although forces will be aware of Algo-care and the ICO

---

[111] See for instance Marion Oswald, 'Algorithmic-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power' (2018) *Phil. Trans. R. Soc. A,* 376:2128.

[112] n76.

[113] https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/12/ico-launches-tool-to-help-police-forces-using-data-analytics/

toolkit to a greater or lesser extent, there is no compulsion to use them. There is also a risk that the deployment of Algo-care and similar frameworks becomes more akin to an 'insurance policy'[114] and less about creating genuine understanding and scrutiny. We have not yet seen legal requirements around AI and data analytics being translated into authorised professional practice or specific codes or policy.

The EU draft AI Act,[115] although not directly applicable to UK policing, proposes a tiered risk-based framework for the regulation of AI with the aim of achieving additional legal certainty, predictive data analytics used by law enforcement being regarded as 'high risk' and therefore subject to ex-ante and ex-post assessments and monitoring. A product safety approach to AI could be criticised as paying insufficient attention to the deployment of AI within a wider decision-making process, the importance of which has been highlighted by the committee's deliberations. Despite this, UK policing will need to pay close to attention to this EU direction of travel to ensure that UK legal, policy and governance frameworks do not fall behind.

Neither are there clear **scientific standards** – written with the policing context in mind – a vital aspect of the second pillar above. Much emphasis is often placed on the positioning of a human police officer in the decision-making loop. However, as Kotsoglou and Oswald query '[h]ow valid is that safeguard in circumstances when an officer on the ground is presented with a finding from Automated Facial Recognition [or other forms of AI] and requested to act on it, especially where decisions around the values built into the tool (and therefore the uncertainties) are made elsewhere?'[116] **Ethical standards** can furthermore contribute to building trustworthy use but only if they link to an individual's personal responsibility, as the police Code of Ethics[117] and the Seven Principles of Public Life (known as the Nolan Principles)[118] attempt to do, and crucially, to policy, processes and audit that enable the standards to be put into practice and to be enforced.

The final pillar of the three-pillar approach is '**people**', both within policing bodies and in associated advisory and oversight functions. This pillar may seem odd in a socio-technical context but I would argue that the success of the overall 3-pillar approach is dependent on the input of people who understand and are committed to the underlying values that the law and ethical standards represent, and are prepared to be thoughtful, engage in professional and skills development and, empowered to constructively question the development and deployment of AI, and whether it should be deployed at all. Again, this is easier said than done, and requires a culture, led from the top, which welcomes informed challenge and transparent oversight and accountability. Transparency combined with

---

[114] Police practitioner interview with author, 2020.
[115] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, COM/2021/206 https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206.
[116] Kyriakos N. Kotsoglou, Marion Oswald 'The long arm of the algorithm? Automated Facial Recognition as evidence and trigger for police intervention' *Forensic Science International: Synergy*, Volume 2, 2020, Pages 86-89, https://doi.org/10.1016/j.fsisyn.2020.01.002.
[117] https://www.college.police.uk/ethics/code-of-ethics.
[118] https://www.gov.uk/government/publications/the-7-principles-of-public-life.

ethical scrutiny can be uncomfortable: 'this ethics oversight has sometimes raised inconvenient concerns that require the data scientists go back to the drawing board'.[119]

Underpinning the three pillars is independent oversight and advice. A national model based on the West Midlands prototype could contribute to the assurance and monitoring anticipated by the draft EU AI Act, the development of necessary policy and proactive longer-term thematic review, provided that the status, resourcing and operational challenges identified above are addressed, and subject to linkage with appropriate regulation and enforcement.

Conclusion

The first three years of voluntary involvement with the committee have been extremely demanding; yet the work has brought home the benefits of productive challenge where independent, critical and even sceptical voices are included. The structure is by no means perfect but the fact that the volume of work for the committee has increased over time is an encouraging sign of the value that transparent independent advice can offer when combined with a process of rolling review. Internationally, the New Zealand Police Service has established a similar external, independent, expert panel to provide advice and oversight from an ethical and policy perspective on emergent technologies[120] and future research might compare and contrast experiences in this regard. In the wider UK, the Scottish Government's independent advisory group on emerging technologies in policing will advise on current legal and ethical frameworks in light of emerging technologies and analytical techniques.[121]

A national ethics approach based on the West Midlands model could, McNeil argues 'help police forces traverse the new technological era and the difficult societal debates that accompany their advancements and implementation. Placing transparency, human rights and diversity of perspective at the heart of policing AI, serves to simultaneously improve the technology, ensure its goals are consistent with a fair society and build public trust and legitimacy.'[122] Yet for such value to be realised at a regional and national level, resources must be found to support the committee's operation on a longer-term, independent basis.

The committee has played a role in addressing the difficult debates that accompany implementation of new technology in policing. A national version has the potential to engage more widely with police forces, civil society and the public, and to further address the legal, policy and ethical issues raised by the increasing emphasis on a 'public health'

---

[119] Tom McNeil 'The emergence of predictive policing – the national issue requiring a national response' *FST Blog*, 9 September 2020 https://www.foundation.org.uk/Blog/2020/The-emergence-of-predictive-policing-%E2%80%93-the-nationa.

[120] https://www.police.govt.nz/about-us/programmes-initiatives/police-use-emergent-technologies.

[121] https://www.gov.scot/publications/independent-advisory-group-on-emerging-technologies-in-policing-purpose-and-remit/. See also Liz Aston 'Second meeting of the Independent Advisory Group on Emerging Technologies in Policing' 24 March 2021 https://blogs.napier.ac.uk/enuapplyscience/2021/03/24/second-meeting-of-the-independent-advisory-group-on-emerging-technologies-in-policing/.

[122] n119.

approach to policing, the implications of which – and the required safeguards - are not yet fully understood. Ultimately however accountability for use of technology by the police remains with the police. An advisory body, however skilled and participatory, cannot be a substitute for a knowledgeable, skilled and experienced stakeholder community. In order to ensure that hard pressed public funds are invested wisely, there needs to be fewer generalisations and more specifics as regards the application of relevant law to the deployment of emerging technologies, recognising the importance of a human rights based approach, combined with agreed scientific standards to help us decide whether things 'work' in the policing context – in which circumstances will they help to understand reality and in which circumstances will they 'come up short'[123] and so should be discounted. Only then will the three-pillar approach recommended in this article be complete and capable of fully supporting trustworthy and accountable use of AI within policing.

---

[123] n96.