

Written Evidence Submitted by Dr Marion Oswald MBE (GAI0012)

Introduction

1. The submission reflects my practical experience and research in respect of the use of AI and data analytics in the public sector, and in particular within policing and criminal justice.
2. This submission sets out my personal views and does not represent the views of Northumbria University, the Alan Turing Institute, RUSI, West Midlands PCC, West Midlands Police, or the CDEI.

Effectiveness of governance of AI

3. In my opinion, there are two key weaknesses in the current governance of the use of AI by the public sector (and in respect of AI procured by the public sector from commercial providers): first, the current lack of regulatory focus on the **use of AI within an operational context** as governed by existing law; and secondly, the importance of a **combined approach to governance of AI**, inter-linking the application of law, technical and ethical standards, empowered and accountable people, and oversight and regulation. By addressing these issues, **innovation will be supported** not hindered, and legal risk reduced. This is because innovators, and those procuring AI, will be given the knowledge and confidence to determine **what is good innovation and what is not** (and so what types of AI should be developed and procured and what should not be).
4. We continue to see many examples of 'AI' promoted for use in public functions, such as 'AI weapons detection' <https://www.bbc.co.uk/news/technology-63476769> or to predict violent behaviour on trains <https://www.dailymail.co.uk/news/article-11422789/CCTV-using-artificial-intelligence-detect-train-sex-pests-thugs-attack.html>, **without robust independent validation of claims made**, and despite the ICO's recent warning in respect of unproven biometric-based AI: <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/10/immature-biometric-technologies-could-be-discriminating-against-people-says-ico-in-warning-to-organisations/>. Many of the ICO's concerns around lack of validity of 'emotion AI' could apply equally to the ongoing use of the **polygraph** within the criminal justice system in England and Wales (Kotsoglou and Oswald, 2021).

Operational and legal context

5. Codes, guidance and regulation that treat AI as a technology in a vacuum, and which use self-evident high-level themes such as 'safety' and 'fairness', will have little or no positive effect, either on confidence to innovate or on legitimacy of use. While useful as a high-level starting point, such principles require considerable **enhancement and expansion** in order to address **complex real-life operational scenarios**. It is crucial that AI is both developed and its use regulated with **its intended or actual operational use in mind**.
6. 'Old' administrative law principles already provide a framework by which the use of AI within public sector decision-making can be considered: in each algorithmic-assisted environment, a **context-specific and nuanced approach will be required so that the information and explanations provided to aid intelligibility, or the way the result is interpreted, enable the particular public task to be fulfilled in a legitimate manner**. (Oswald, 2018)
7. Furthermore, AI tools are not deployed in a vacuum but become part of an operational decision-making system. Therefore, the **laws and regulations that govern that particular decision apply to the use of AI within that decision-making context**. For example, police powers of stop-and-search and arrest require **objective grounds for reasonable suspicion** based on facts, information and/or intelligence.

8. Codes and guidance have yet to address adequately the use of facial recognition technology (for instance) in these specific contexts i.e. addressing if and in what circumstances/under what conditions the output of a facial recognition tool can be treated as objective grounds for reasonable suspicion justifying a stop-and-search or arrest. The answer to this question requires **detailed and specific understanding of the probabilistic nature of AI and how it works**, including consideration of issues such as error rates, sensitivity settings, likelihood of bias and the circumstances of deployment. ('Reasonable' appears elsewhere in the law as an important concept, yet is not defined in AI terms, thus emphasising the need for interpretation and guidance.)

9. In other words, **we need to know if the AI is any good for the specific context in which it will be deployed and the legal test which needs to be satisfied**. These determinations should then feed back into the technical development process, thus improving quality. Claims made by manufacturers about a technology's ability to solve a problem (such as detecting all weapons, identifying 'risky' individuals or making evidence collection more effective) should not be accepted at face value.

10. Another context which illustrates these challenges is the concerted effort in recent years to embed a 'public health' approach in criminal justice, underpinned by Government priorities focused on risk, harm prevention/reduction, vulnerabilities, and youth and domestic violence. A key element is the skilled use of data science techniques, particularly at a population level and for individualised risk assessment and prediction.

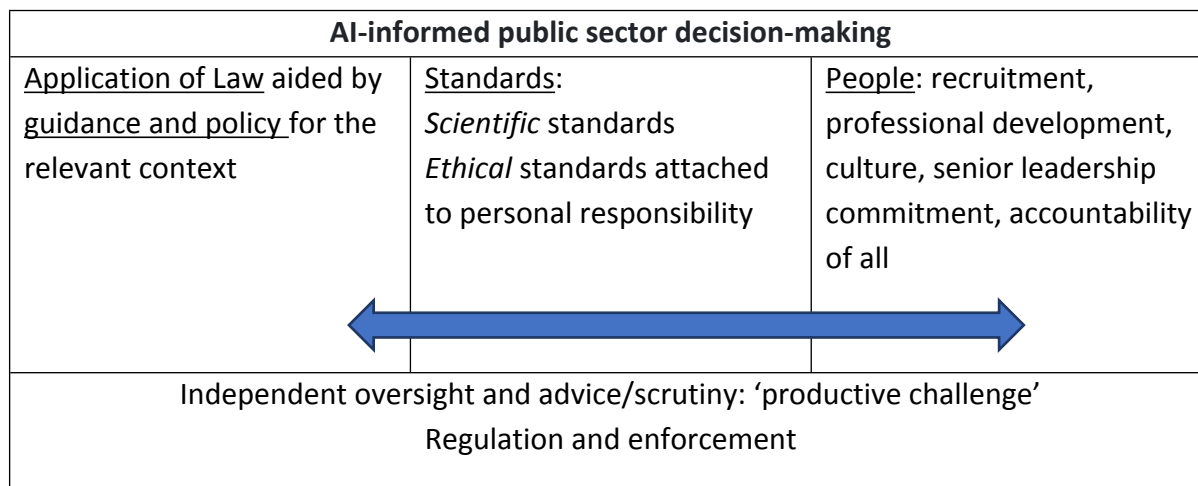
11. However, the models of preventative actions and the consequences of acting on a **'false positive'** AI output will be significantly different in a public health context as opposed to a criminal justice one. In a health context, a person may be subjected to an additional test or examination to re-check their health condition or may be given unnecessary treatment. In a criminal justice context, a person may be recorded in a police system in a high-risk category, thus affecting the accuracy of those records and how that person is treated in the future, or the person may be subjected to additional surveillance, investigation or police intervention. There is therefore no one-size-fits-all answer to whether it would be legal, ethical and effective to incorporate a particular AI tool in these contexts, despite the 'public health' badging of the aims.

Review, scrutiny and oversight

12. **Technical, statistical, legal, contextual, operational and ethical aspects** of AI-informed decision-making are closely interconnected (Oswald, 2022). We need to know **what the output means in the context of the operational decision to be taken, and we must assess the implications of how it will be used in practice**.

13. Take a hypothetical example of a public sector body which has been given a power to intervene 'if a child is reasonably determined to be at high risk of harm'. The legality of its power to intervene is therefore dependent upon this assessment (which can and often is tested in court). The body might decide to use an algorithm to help it decide on risk levels by way of analysis of hospital admission reports. But unbeknown to users, the model both has a **high level of false positives** and is based on machine learning textual analysis of terms (chosen by the commercial developer) that expert users would regard as **irrelevant** to their assessment of the risk (and missing factors in other records that users would regard as relevant). **So the tool is not in fact answering the question that needs to be answered by the human decision-maker and is based on irrelevant factors**. If the body defers to the tool and therefore intervenes in respect of a child assessed wrongly due to issues with the tool, then it will be acting outside its powers (not to mention reducing services to children at real risk of harm and causing unnecessary disruption, stress and embarrassment to the child and family).

14. The three-pillar approach set out below attempts to illustrate the inter-linking of all these factors and is further discussed [here](#):



Practical implementation

15. There are a number of practical implementation challenges raised by the above three-pillar approach. First, the **application of relevant law** is not easy, as it often requires the 'read-across' of common law principles into new contexts and the translation of key legal requirements into **suitably precise policy and guidance**. Secondly, **scientific and ethical standards** must address the specific operational challenges that will be encountered in the particular context. More attention is therefore needed to the acquisition of '**off-the-shelf**' commercial applications or models created through the combination of a number of tools, none of which have been specifically developed or evaluated for the context in question. Thirdly, organisations require a culture, led from the top, which welcomes informed challenge, supported by **empowered staff** who understand and are committed to the underlying values that the law and ethical standards represent, and are prepared to be thoughtful, and engage in professional skills development (therefore requiring bodies such as the College of Policing to provide the means for such professional development).

16. It will be crucial, as the AI environment becomes ever more complex, that **independent scrutiny and regulation** is '**end-to-end**' i.e. that scrutiny - and thus accountability - becomes a **rolling process**, from project planning/initiation to eventual operationalisation, rather than being limited to ex-post review. New models of review and scrutiny will be needed, and lessons could be learned from the proceedings of the **West Midlands Police and Crime Commissioner and West Midlands Police data ethics committee** (the first of its kind in UK policing) which is an ongoing experiment in scrutinising and advising upon AI policing projects proposed for real operational environments. National models based on the West Midlands prototype – provided properly resourced and financed - could contribute to the assurance and monitoring required, the development of necessary policy and proactive longer-term thematic review.

Transparency

17. The Government's draft [Algorithmic Transparency Standard](#), which has recently been piloted with a number of public sector bodies, shows potential to improve the transparency of algorithms used in the public sector, provided that i) its **adoption** by the public sector can be secured, ii) the final format accommodates the needs and interests of **different stakeholders** (an informed but non-expert public, civil society, journalists, technical experts etc), and iii) the required **resources** are allocated to adoption and long-term maintenance. It will also be crucial that **supplier responsibilities** to support compliance with transparency – including appropriate disclosure of algorithmic functionality, data inputs and performance - are covered in procurement contracts and addressed up front as a mandatory requirement of doing business with the public sector.

18. My recent research into the Algorithmic Transparency Standard and Police Perspectives ([Oswald et al., 2022](#)) concluded that, despite the sensitive nature of some policing AI, the rewards for the police of a carefully tailored Standard implemented at the right stage of algorithmic development outweighed the risks. Participation in the Standard provides an opportunity for the police to demonstrate the **legitimacy** of technology use and build earned trust.

19. Furthermore, the Standard could also be used to develop **increased sharing** among police forces of best practices (and things to avoid). Research participants were keen for compliance with the Standard to become part of a system to **drive reflective practice** across policing around the development and deployment of algorithmic technology. This could enable police to learn from each other, facilitate good policy choices and decrease wasted costs. In order to contribute to improving the quality of policing technology, the Standard should be **linked to methods of oversight and promotion of best practice** on a national basis. Otherwise, the Standard may come to be regarded as an administrative burden rather than a benefit for policing, a concern that would apply across the wider public sector.

References:

Marion Oswald and Chambers, Luke and Goodman, Ellen P. and Ugwu-dike, Pam and Zilka, Miri, The UK Algorithmic Transparency Standard: A Qualitative Analysis of Police Perspectives (July 7, 2022). Available at SSRN: <https://dx.doi.org/10.2139/ssrn.4155549>

Marion Oswald, 'A Three-Pillar Approach to Achieving Trustworthy Use of AI and Emerging Technology in Policing in England and Wales: Lessons From the West Midlands Model' (2022) *European Journal of Law and Technology* Vol. 13 No. 1 <https://ejlt.org/index.php/ejlt/article/view/883>

Kyriakos N. Kotsoglou and Marion Oswald, 'Not 'very English' - On the Use of the Polygraph by the Penal System in England and Wales' (2021) *Journal of Criminal Law* 85(3) 189-208
<https://journals.sagepub.com/doi/full/10.1177/0022018320976284>.

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
<https://royalsocietypublishing.org/doi/full/10.1098/rsta.2017.0359>.

(17 November 2022)