Ethical review to support Responsible Artificial Intelligence (AI) in policing:

A preliminary study of West Midlands Police's specialist data ethics review committee

Executive Summary, Conclusions and Recommendations

Marion Oswald,[1] Claire Paterson-Young, Pauline McBride, Michael Maher, Muffy Calder, Gitanjali Gill, Elizabeth Tiarks and William Noble

11 September 2024

The project is part of the BRAID (Bridging Responsible AI Divides) Programme and funded by the Arts and Humanities Research Council (AHRC), part of UK Research & Innovation (UKRI)

[1] **Corresponding author and project lead** marion.oswald@northumbria.ac.uk

Executive Summary

The West Midlands Police and Crime Commissioner (WMOPCC) and the West Midlands Police (WMP) have for the past five years maintained a Data Ethics Committee to advise on the design, development and deployment of advanced data analytics and AI capabilities. This Committee comprises people drawn from backgrounds in academia, industry, public/third sector and policing. Since 2019, it has met at least on a quarterly basis, advising and making recommendations on each occasion on several projects and proposed tools, from in-principle analysis to tools ready for operational use. Its papers and minutes are published via WMOPCC. This interdisciplinary research used mixed-methods (including 26 interviews) to review the **impact and influence** of the Committee, and to recommend to national bodies, other forces and to WMP/WMOPCC factors that affect how best to go about using independent advisors in this context. Lessons from the Committee's experience, together with a single structured framework could inform a **coherent and consistent national approach**. The Conclusions and Recommendations (for **national strategy, police, Committee members, community representatives, academia and research funding bodies**) fall into the following themes:

- A Data Ethics Committee with diverse independent voices can **contribute positively** to the validity and responsibility of policing AI, thus **supporting operational policing**. It can develop **understanding** within the police of key ethical, scientific, legal and operational issues for planning and implementation. This will be successful only if the Committee has a clear function, is **fully incorporated** into the system of oversight and scrutiny, visibly championed by the Chief Constable & PCC, and suitably supported by a secretariat, robust process and communications;

- This will be successful only if membership includes genuine representation from the **community** that the police serves, there is transparent engagement, and time taken to allow members to understand the **technical and legal** aspects of the work.

- This will be successful only if the **operational context** is explained by operational police officers, and time taken to understand **how AI outputs will be used**, so as to enable potential benefits, risks/harms and proportionality to be assessed in the same conversation. Attention must be paid to police responsibilities for **public safety** (and how AI may support these responsibilities) as well as to risks related to **privacy, fair trial and freedom of expression**.

- Police forces, PCCs and national bodies embarking on such an approach will need to be prepared for **ambiguity**. There are often **no 'black and white' answers** to ethical, legal or technical questions raised by policing AI, such as reconciling privacy and security priorities relevant to the assessment of the proportionality of using suspect data.

Conclusions

Research Question 1- *Influence on technology*: What influence has the Committee had on the design and operationalisation of WMP AI policing projects?

The research found that the Committee exerts **influence**, though steps could be taken to improve transparency around the extent of that influence.

The research found that the Committee's exchanges with the Data Analytics Lab influences the design, operationalisation, transparency and good practice of AI projects in policing. However, there are challenges in terms of the time and expertise required for Committee members to understand the **technical detail** of the models (such as feature engineering) and thus the implications of the outputs.

The exchange between the Committee, WMP and the Data Analytics Lab influences the design and operationalisation of AI projects in more subtle ways. The Data Analytics Lab **anticipates** the concerns of the Committee in its approach to new projects. Senior police officers responsible for the delivery of projects speak of **insights** they have gained from these exchanges. The influence of the Committee extends beyond WMP when these officers join other policing bodies or take on national portfolios and strategic roles.

The extent to which the Committee's recommendations are followed is not immediately clear from a bare reading of the minutes. In the interests of transparency, informing public debate and improving trust, records should be maintained that not only detail the projects considered by the Committee and the Committee's advice and recommendations, but also the **impact** of those advice and recommendations on the development of the project. Where appropriate, these records should be made accessible to the public.

Research Question 2- *Human rights issues*: What human rights related issues were identified by the Committee and how were these issues dealt with in the design and operationalisation of AI tools?

The research found that the Committee has a strong human rights focus. Data Analytics Lab and Police representatives described how engagement with the Committee **enhanced or transformed** their understanding of issues with rights implications and informed their own approach, and that of the Data Analytics Lab, to the development of projects.

However, the research highlighted concerns about the productiveness of the Committee's human rights discourse, and that it may be **overly focused** on human rights relating to privacy and protection from discrimination. Interviewees, including Community representatives, spoke of the need to broaden rights-based conversations to take account of other rights, mentioning the right to a **fair trial** and the **state's positive obligations** under the prohibition against torture and right to life.

Committee representatives also expressed concerns about lack of information which would allow them to assess the **real-world outcomes** resulting from the operationalisation of AI projects. Lack of information about outcomes negatively impacts on the Committee's ability to assess the human rights and ethical implications of deployment of AI projects, and to understand technical successes and challenges.

Interviewees noted that rights issues are frequently raised without the Committee explicitly mentioning the right(s) at stake. This makes it difficult to identify **which human rights issues** are raised by the Committee from a reading of the minutes and impacts on the utility of the minutes as a source of learning and aid to transparency. It would be helpful for the minutes to explicitly identify which human rights issues were raised by the Committee.

Interviewees expressed concern about the lack of clarity as to when projects would **return to the Committee** for discussion and advice after operationalisation. This impacts on the ability of the Committee to fulfil its remit and address human rights and other ethical issues that may only become apparent after operationalisation.

Research Question 3- *Vulnerable groups and data*: How, if at all, are the interests, views and concerns of vulnerable groups incorporated within the ethical review process?

There are significant barriers to improving the representation of vulnerable communities within the Committee. It should be noted, however, that the Committee considers the interests of the community in its **discussions around privacy, disproportionality, and safeguarding**. The research identifies three stages of community representation that need to be addressed:

- First, there is **Accessibility and Definition**. An early barrier to community engagement is that community representatives are unsure of the role of the Committee within the policing ecosystem and how it might positively impact on the community. Community representatives are more familiar with the role of scrutiny panels, for example, and see involvement in those as a more effective use of their time. It is, therefore, important to promote the work of the Committee to potential Community representatives.
- Secondly, **Capacity and Influence** should be considered. It should be communicated to potential Community representatives that their value is in their contextual and personal experiences. Community representatives discussed feeling that a lack of technological knowledge would restrict their ability to engage in discussions. However, the other stakeholder groups did not consider this to be an issue, and that contextual conversations should be encouraged.
- Finally, **Dissemination and Development** should be employed to strengthen engagement with vulnerable groups. Direct engagement with communities is essential for better representation and the inclusion of other vulnerable voices within the community. The research suggests that it may be beneficial for Community representatives who join the Committee to act as community advocates, disseminating information about the Committee to the wider community, though other approaches may also be effective.

Research Question 4- *Challenges of ethical review*: What issues and challenges have Committee members and police representatives encountered in the committee review process?

In large part, the Committee fulfils its mandate and provides effective scrutiny of AI policing technologies. There are challenges, however.

There is a need to re-evaluate **communication processes** to ensure that the Committee has visibility as to whether, for each project, its recommendations have been followed. Consideration should be given to supplementing the Committee's recommendations to the Data Analytics Lab with an explanation of the **rationale** for the recommendations. This would allow the Data Analytics Lab to contextualise the Committee's concerns and respond appropriately. Currently, there are indications that absence of this information negatively impacts the efficiency of the review process.

There is a need for a clear protocol for the **return of 'live' projects to the Committee** for consideration and review post-implementation. The protocol should set out the timescales and other 'triggers' for the return of projects to the Committee. This would mean that active projects with real-world impacts could be fed into the decision-making process for future recommendations.

The use of **technical language** was identified as a potential barrier to involvement of community representatives and may slow down discussions. Consideration should be given to finding additional time and space during meetings for the technical discussions.

Although the Committee's remit clearly requires it to consider the impact of AI projects intended for use in policing, some Data Analytics Lab and Police representatives expressed concern that the Committee occasionally sought to challenge the **operational independence** of the police. The high turnover of police workforce may impact on the ethical review process through loss of established relationships. This may also impact on the work of the Data Analytics Lab as focuses and priorities change. Both these issues may be mitigated by the regular and consistent involvement of **senior operational police officers** in the Committee meetings in order to ensure that the operational priorities and actions behind the AI tools are explained and understood.

Research Question 5- *Potential of other models to improve the Committee process*: In what ways could the use of the factor's framework (Janjeva, Calder and Oswald 2023) and matrix evaluation model (Oswald, Chambers and Paul 2023) improve the development of Responsible AI in policing?

The research suggested that a structured framework for evaluation of AI projects would help maintain accountability, enable a detailed assessment of the impact of a project on human rights, be educational, and could form the **foundation stones** for ethical review. It may also support the integration of broader rights conversations, as specified in Research Question 2. The research's technical observations indicated how user information that indicates **probability and certainty** of model outputs to the police user could reduce the risk of overreliance. Further research and testing are necessary to explore the potential for the factors framework (Janjeva, Calder and Oswald 2023) and matrix evaluation model (Oswald, Chambers and Paul 2023) to improve the development of Responsible AI in policing.

Research Question 6- *Research challenges*: What challenges emerge from the research which would need to be addressed in larger research projects investigating embedded ethics processes?

The key challenges that emerged during this research were as follows:

- There was a **low response rate** from community members for interviews, with some of those responding, expressing interest but unable to commit due to other obligations. In future, alternative research methods, such as questionnaires, should be considered to complement interviews and provide avenues for individuals in demanding and resource limited sectors to participate in the research.
- The **thematic analysis** of the meeting minutes was challenging due to the limited detail included in the minutes. Implementing standardisation and clearer communication guidelines could make this approach more sustainable in future research. This would not

only aid future research endeavours but would provide a structured feedback mechanism that could improve the Committee's review and feedback.

- Technical observations were possible through the co-operation of the Data Analytics Lab and the coordination of staff from the WMOPCC. It would have been beneficial to observe operational systems with 'live' data, but it would not have been practicable to secure **vetting** for the researchers within the timeframe of this project. This demonstrates wider issues which would need to be addressed in larger research projects investigating embedded data/AI ethics processes in policing.

Recommendations

For national and international strategy, policy-makers and stakeholders

This research indicates that development of policing AI that includes consultation and advice from diverse independent voices can contribute to the robustness of technology implementation, from technical, human rights, operational and community perspectives. For such advice to add value, however, it must be **incorporated fully** into the implementation and oversight processes, and not regarded as an 'add-on' or tick-box exercise.

Although it takes time and effort to construct, embed and refine such an advisory process, it can contribute to a positive culture whereby the police develop **knowledge and understanding** of the issues likely to be raised by such independent oversight, thus enabling those issues to be anticipated and considered in **planning and implementation**. There can be tension between Committee members and police staff regarding the extent to which operational decisions fall within the Committee's remit. Committee members assert however that **deployment in practice of AI outputs must be understood, to assess potential benefits, risks/harms and proportionality**. The regular involvement of **operational police officers** to discuss and explain operational priorities and actions behind the AI tools can contribute to resolving these tensions. Furthermore, attention must be paid to police responsibilities for **public safety (and how AI may support these responsibilities)** as well as to risks related to privacy, fair trial and freedom of expression.

Having a data ethics process cannot be assumed to create or improve trust in policing AI per se, especially from vulnerable groups. For this to happen, effort must be put into incorporating the **voices of the community** into the process and ensuring that the work of the committee/panel is known and that its input is respected and influential.

Steps should be considered to highlight and disseminate information about the work of the Committee and the Data Analytics Lab in a manner that allows other police authorities, ethics committees, policy-makers, oversight bodies and international stakeholders to benefit. Lessons from the Committee's experience, including those reviewed in this report, can inform **best**

practice and feed into a framework for responsible AI in policing, including a **national model** for independent advice and oversight.

Such lessons include consideration of whether review of projects fully addresses all those human rights issues that are at stake, including the **positive rights** under Articles 2 and 3 (respectively, the right to life and freedom from torture/inhuman and degrading treatment). It should be noted that the Committee's experience shows that there are often **no 'black and white' answers** to ethical, legal or technical questions and the key issue of proportionality (e.g. reconciling privacy and security priorities relevant to the assessment of the proportionality of using suspect data, and whether precision or other performance metrics should be prioritised given the nature of the project). This research indicates however that a **structured framework** such as that proposed by Janjeva et al. (2023) might improve the productiveness, robustness and objectivity of deliberations about necessity and proportionality and assist in the practical application of the proportionality test when dealing with technical issues of data analytics and AI.

Building a culture of Responsible AI in policing depends on **time, resource, commitment, knowledge and collaborative communication**. It is important for police authorities to be aware of the potential issues that may be raised by independent oversight bodies, so they can plan and prepare for those conversations, while those involved in the oversight must be aware of the operational purposes and objectives of AI projects. This requires an **openness to both teach and learn** from other groups, and investment of time and resource in relationship-building.

For the WMOPCC, the Secretariat, WMP and Committee members

Naming and role of the Committee/Terms of Reference: consideration could be given to the renaming of the Committee in order to better reflect and communicate its advisory role e.g. Advisory Committee. Furthermore, it is recommended that the Terms of Reference are reviewed in light of any accepted actions from this report.

Performance metrics: there is an ongoing need for Committee members to understand the significance of the range of performance metrics (accuracy, sensitivity, precision, specificity) for predictive models, how these metrics are assessed, why the Data Analytics Lab may have favoured one metric over another, when one metric should be preferred over another and what

that selection might mean for the impact of the tool and the circumstances in which/purposes for which the tool should be used.

Communication Strategy/Process Map: Clearer protocols should be developed for the return of 'live' projects to the Committee to improve and inform future discussions, with triggers for return agreed on a case-by-case basis. This should be built into a clear process map that informs stakeholders about how projects can be returned to the Committee, and what they should expect in terms of feedback, recommendations, and advice. The Committee, Data Analytics Lab, WMP and WMOPCC should collaboratively discuss what information needs to be shared, when, by and with whom, about the intended purpose and use of projects and the likely impacts, whether for the purpose of ethical review or to guide the development of the project.

Project Tracking: In the interests of transparency, informing public debate and improving trust, records should be maintained that not only detail the projects considered by the Committee and the advice and recommendations, but also record the impact of the advice and recommendations on the development of the project. Where appropriate, these records should be made public.

Precedent bank: The value of keeping a record of outcomes throughout the lifecycle of projects was noted throughout the research. Building a 'precedent' bank of projects reviewed by the Committee could develop knowledge of technical, operational and ethical issues raised, and their solutions/mitigations, thus making future decision-making in relation to similar projects more effective.

The Minutes: The recommendations set out in the minutes should be supplemented with an explanation of the rationale for the Committee's recommendations including statements on the human rights and other key technical, legal and social issues identified by the Committee. This would allow the Data Analytics Lab to contextualise the Committee's concerns and respond appropriately. In order to increase communications and understanding, it is recommended that time be built into the meeting agendas to allow the DAL to return to hear the recommendations and rationales for these.

Annual Report: Recognising that additional record-keeping requires additional resources, it is recommended that in the interests of transparency and as a means of informing public debate and improving public trust, the Committee should produce an annual report which summarises

its work and offers a brief commentary on human rights and other key issues addressed by its recommendations. This would additionally support public knowledge of its role and remit.

Impact of Technical Language: Both Community representatives and Committee members noted that the use of technical language can act as a barrier to participation and impede discussion at meetings. To address this, stakeholders should consider solutions, such as the timetabling of additional time for technical discussion, or members given specific responsibilities regarding the explanation of technical details to other members.

For the WMOPCC, Community representatives and Committee members

Enhanced inclusion of Community representatives: Consideration should be given to developing an engagement plan for community members to ensure they are represented in the Committee and are able to contribute appropriately. Further to this, opportunities should be sought out to improve public knowledge of the Committee, its remit, objectives, and potential benefits. As part of this, direct community engagement should be considered through practices such as informal meetings, training sessions, workshops, or engagement panels.

Creating Space for Community Voices: The research found that a concern of community representatives was that they felt they could not contribute to discussion on AI or data ethics, and that a lack of knowledge would leave them 'exposed' during meetings. Although Committee members recognised the value of community voices, and the contextualisation of ethical concerns, the cultivation of inclusive conversations and welcoming environments should be prioritised to help develop community confidence.

For academic researchers, and research funding bodies

Proportionality and policing AI: Assessment of proportionality and human rights-related risks concerning the use of policing AI in real contexts is not straightforward and can be fraught with ambiguity, particularly where law enforcement bodies may be expected to use new methods of data analytics and AI to tackle public safety risks. It is clear from the Committee's experience that a proportionality assessment that is alive to how the technical, statistical and operational details of policing AI link to the legal test is dependent on an **interdisciplinary approach**. More research

could be done to consider how the proportionality test could be better applied in 'messy' realities, and to avoid proportionality merely becoming a '***ritual incantation***'.

Challenges of AI-related research in sensitive contexts: considerable issues of **access** arise in research projects investigating data/AI and ethics processes in policing, even for researchers with existing research relationships and who hold levels of vetting. Funding bodies and national research programmes could consider ways that appropriate and systematic access could be given to researchers on a more equitable basis. However, building research relationships and knowledge of the consequences of AI in this context requires time and commitment and therefore **longer cycles of funding** would be more beneficial in this regard.

Research limitations

The researchers acknowledge a number of limitations and constraints. Two of the researchers involved in the project are members of the Committee, which could have introduced an element of bias. These researchers were primarily involved in the direction, oversight and management of the project. However, one researcher was involved in elements of data collection and recruitment as well as reflective analysis sessions to identify and refine themes for the research, and one researcher was involved substantially in the write-up and finalisation of the report. In order to reduce any potential bias in selection, interviewing and transfer, the following processes were employed:

- Embedding reflexivity in the approach to research.
- Participant selection of Committee and DAL representatives was completed by members of the research team who had no direct link with the Committee.
- Clearly defined topics were introduced to the interviews to standardise the process.
- Analysis of interviews with Committee and/or DAL representatives was completed by members of the research team who had no direct link with the Committee.

In addition, the representation of community voices was limited, primarily coming from advocacy groups that promote social inclusion and equality. This may have limited an intersectional approach to the research (see: 'Challenges to the intersectional research approach' section in the main report). For future studies, we recommend including perspectives from general community groups.

The number of systems included in the technical observations was limited. Access to a greater spread of AI systems would have provided us with a better sense of how police officers and other users interact with AI systems in use in a policing context, what information is communicated via such systems about their reliability, capabilities and limitations, and what training if any is provided about the systems.