Expert Panel on Emergent Technologies
Advice on Taylor & Fry report for NZ Police: Safe and ethical use of algorithms (EPET 21.1)

21 May 2021

1. The panel welcomes the opportunity to review the above report and provide feedback to the Police.

2. The terms of the referral were as follows:

> the Panel's advice is sought on implementing the report's recommendations including identifying te ao Māori and broader ethical considerations, and identifying any areas that may benefit from ongoing Panel involvement during the implementation process.
>
> Specific advice is not sought on the scope or methodology adopted by Taylor Fry in responding to its engagement, although the Panel may wish to offer general observations on additional/alternative considerations that ought to inform judgments about the risks posed by different types of algorithms.

3. In general terms, the Panel considered the report a useful starting point for a discussion on ethical use of algorithms, and a helpful source of information. Most of the recommendations are of a kind that we would broadly expect and support. The general principles listed at [3.2] include the sorts of considerations that should precede any decision to proceed further with AI development or procurement, most notably "What problem are you trying to address?" and "Is it appropriate to use an algorithm for this?"

4. The adequacy of the more substantive proposals, such as that "a proper governance framework", "approval, monitoring and ongoing review processes" and a "formal evaluation structure for algorithm developers" should be put in place, will depend substantially on details not included in the report. In particular, while we strongly endorse these recommendations, much work remains to be done on the specifics of the necessary processes and structures, and the criteria against which algorithms will be assessed. This is also true of the formal monitoring and auditing processes. Such steps are indispensable to the safe and ethical deployment of any algorithms, but the devil lies very much in the details. The adequacy of NZ Police's algorithm policy will depend on how such mid-level commitments are designed, implemented and monitored and evaluated.

5. As a comprehensive overview of ethical considerations around the ethical use of algorithms, the panel identified certain shortcomings. Most obvious among these was the absence of any specific reference to te ao Māori or commitments under the Treaty of Waitangi, or indeed, any recognition of any considerations distinct to Aotearoa New Zealand. The Panel recommends that upstream engagement with Māori and other communities should be prioritised, and serious attention paid to

active participatory co-design, rather than consultation at a later stage in the algorithm's development when the engagement may be in the form of 'take it or leave it'.

6. An important consideration about any governance process relates to the initial classification of algorithms as high, medium or low risk. Such a classification system is common to many governance processes, including NZ's Algorithm Charter and the recently published EU Commission's draft Regulation on AI.[1] Since this early assessment will substantially determine the degree of scrutiny to which they are subsequently subjected, the Panel wishes to identify this step as a potential weak link in any governance scheme.

   As the Taylor and Fry report notes, "allocating algorithms into high, moderate and low-risk categories is somewhat arbitrary and subjective." Proper attention must therefore be paid to the criteria for this initial classification, and in particular, to ensure that the criteria not only include technical and data issues, but also assess the implications of the individual or place-based interventions that may result from use of the algorithm. Moreover, proper attention should be paid to who is involved in this initial classification, as there is a risk that algorithms will be classified as low risk without consultation with those most at risk . Specifically, the Panel expressed some reservations about some of the algorithms listed in Appendix A as "low risk", and in particular, about the first item on that list.

7. In common with many similar reports, Taylor and Fry place significance on "the need to ensure all tools were subject to human oversight." While there is much to commend such an approach, a growing body of literature in this area has drawn attention to the risk that nominal human oversight can offer false reassurance. In particular, concerns about "the control problem"[2,3] would need to be addressed if human control is to be meaningful, which would suggest particular attention needs to be given to the design of the human interface and data visualisation.

8. The Panel were somewhat sceptical of Recommendation 8: 'Develop algorithms nationally, rather than at a district level.' It was not obvious to us that it will invariably be preferable to develop algorithms nationally rather than locally. Local

---

[1] EU Commission. *Europe fit for the Digital Age: Artificial Intelligence*, 21 April 2021. https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_21_1682/IP_21_1682_EN.pdf.

[2] This has been described as "the tendency to over-rely on automated outputs and discount other correct and relevant information." Babuta, A. and Oswald, M. (2019) Briefing paper: Data Analytics and Algorithmic Bias in Policing, at p.15. https://rusi.org/sites/default/files/20190916_data_analytics_and_algorithmic_bias_in_policing_web.pdf,

[3] "'The control problem' arises from the tendency of the human agent within a human-machine control loop to become complacent, over-reliant, or overtrusting when faced with the outputs of a reliable autonomous system." John Zerilli, et al. *A Citizen's Guide to Artificial Intelligence* (MIT Press, 2021), p.83. See also. Parasuraman, R. and Manzey, D.H. (2010) Complacency and bias in human use of automation: An attentional integration. *Human Factors* 52(3): 381-410; Skitka, L.J., Mosier, K. and Burdick, M. D. (2000) Accountability and Automation Bias. *International Journal of Human-Computer Studies* 52: 701-717; Zerilli, J., Knott, A., Maclaurin, J. and Gavaghan, C. (2019) Algorithmic Decision-Making and the Control Problem. *Minds and Machines* 29: 555–578.

development can in some cases pick up regional nuances and issues relating to data quality, gaps and uncertainties.

9. Although our remit was to advise on the report's recommendations rather than the algorithms themselves, we wish to alert you to the fact that the panel expressed significant concern about the road policing algorithm (2.2.3). This seems to us to contain bias reinforcement loops that lie behind many of the most serious concerns about algorithms in policing, and did not appear (from the description in the report) to have engaged fully with relevant legal framework governing a decision to take coercive, intrusive or enforcement action. The Panel would be happy to discuss further.

10. Whilst not being asked to discuss the Taylor Fry methodology, the Panel would like to alert you to the following methodological issue. There is a different ethical framework associated with the conducting of trials or experimental testing of algorithms on the one hand, and the operational application of those algorithm on the other hand. Evaluative research requires the consent of participants and consideration of any potential law breaking during the trial as well as incidental matters that may come to light as a consequence. Once operational, there will be a requirement to have an operational protocol and transparency about the processes and procedures accompanying them.

11. The Panel would like to suggest that the report has perhaps underplayed the issue of scientific validity of algorithmic approaches, particularly those involving individual predictions based upon police data or visualisations/data analysis that could disguise the nuanced, partial or uncertain nature of the underlying data. The Panel would suggest that these issues should be addressed more fully in the evaluation process and criteria.

12. With regard to the possibility of ongoing Panel involvement during the implementation process, we would certainly be open to discussing this further, but our capacity to engage with this process depends substantially on the volume of other referrals we can expect.